Graduate Texts in Mathematics

GTM

John Voight

Quaternion Algebras

OPEN ACCESS

Springer

Graduate Texts in Mathematics 288

Graduate Texts in Mathematics

Series Editors

Sheldon Axler
San Francisco State University, San Francisco, CA, USA

Kenneth Ribet
University of California, Berkeley, CA, USA

Advisory Board

Alejandro Adem, *University of British Columbia*
David Eisenbud, *University of California, Berkeley & MSRI*
Brian C. Hall, *University of Notre Dame*
Patricia Hersh, *University of Oregon*
J. F. Jardine, *University of Western Ontario*
Jeffrey C. Lagarias, *University of Michigan*
Eugenia Malinnikova, *Stanford University*
Ken Ono, *University of Virginia*
Jeremy Quastel, *University of Toronto*
Barry Simon, *California Institute of Technology*
Ravi Vakil, *Stanford University*
Steven H. Weintraub, *Lehigh University*
Melanie Matchett Wood, *Harvard University*

Graduate Texts in Mathematics bridge the gap between passive study and creative understanding, offering graduate-level introductions to advanced topics in mathematics. The volumes are carefully written as teaching aids and highlight characteristic features of the theory. Although these books are frequently used as textbooks in graduate courses, they are also suitable for individual study.

More information about this series at http://www.springer.com/series/136

John Voight

Quaternion Algebras

Springer

John Voight
Department of Mathematics
Dartmouth College
Hanover, NH, USA

This book is an open access publication. The Open Access publication of this book was made possible in part by generous support from Dartmouth College via Faculty Research and Professional Development funds, a John M. Manley Huntington Award, and Dartmouth Library Open Access Publishing.

ISSN 0072-5285 ISSN 2197-5612 (electronic)
Graduate Texts in Mathematics
ISBN 978-3-030-56692-0 ISBN 978-3-030-56694-4 (eBook)
https://doi.org/10.1007/978-3-030-56694-4

Mathematics Subject Classification: 11E12, 11F06, 11R52, 11S45, 16H05, 16U60, 20H10

© The Editor(s) (if applicable) and The Author(s) 2021. This book is an open access publication.
Open Access This book is licensed under the terms of the Creative Commons Attribution-NonCommercial 4.0 International License (http://creativecommons.org/licenses/by-nc/4.0/), which permits any noncommercial use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license and indicate if changes were made.
The images or other third party material in this book are included in the book's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the book's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.
This work is subject to copyright. All commercial rights are reserved by the author(s), whether the whole or part of the material is concerned, specifically the rights of translation, reprinting, reuse of illustrations, recitation, broadcasting, reproduction on microfilms or in any other physical way, and transmission or information storage and retrieval, electronic adaptation, computer software, or by similar or dissimilar methodology now known or hereafter developed. Regarding these commercial rights a non-exclusive license has been granted to the publisher.
The use of general descriptive names, registered names, trademarks, service marks, etc. in this publication does not imply, even in the absence of a specific statement, that such names are exempt from the relevant protective laws and regulations and therefore free for general use.
The publisher, the authors and the editors are safe to assume that the advice and information in this book are believed to be true and accurate at the date of publication. Neither the publisher nor the authors or the editors give a warranty, expressed or implied, with respect to the material contained herein or for any errors or omissions that may have been made. The publisher remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

This Springer imprint is published by the registered company Springer Nature Switzerland AG
The registered company address is: Gewerbestrasse 11, 6330 Cham, Switzerland

Preface

Goal

Quaternion algebras sit prominently at the intersection of many mathematical subjects. They capture essential features of noncommutative ring theory, number theory, K-theory, group theory, geometric topology, Lie theory, functions of a complex variable, spectral theory of Riemannian manifolds, arithmetic geometry, representation theory, the Langlands program—and the list goes on. Quaternion algebras are especially fruitful to study because they often reflect some of the general aspects of these subjects, while at the same time they remain amenable to concrete argumentation. Moreover, quaternions often encapsulate unique features that are absent from the general theory (even as they provide motivation for it).

With this in mind, the main goal in writing this text is to introduce a large subset of the above topics to graduate students interested in algebra, geometry, and number theory. To get the most out of reading this text, readers will likely want to have been exposed to some algebraic number theory, commutative algebra (e.g., module theory, localization, and tensor products), as well as the fundamentals of linear algebra, topology, and complex analysis. For certain sections, further experience with objects in differential geometry or arithmetic geometry (e.g., Riemannian manifolds and elliptic curves), may be useful. With these prerequisites in mind, I have endeavored to present the material in the simplest, motivated version—full of rich interconnections and illustrative examples—so even if the reader is missing a piece of background, it can be quickly filled in.

Unfortunately, this text only scratches the surface of most of the topics covered in the book! In particular, some appearances of quaternion algebras in arithmetic geometry that are dear to me are absent, as they would substantially extend the length and scope of this already long book. I hope that the presentation herein will serve as a foundation upon which a detailed and more specialized treatment of these topics will be possible.

I have tried to maximize exposition of ideas and minimize technicality: sometimes I allow a quick and dirty proof, but sometimes the "right level of generality" (where things can be seen most clearly) is pretty abstract. So my efforts have resulted in a level of exposition that is occasionally uneven jumping between sections. I consider this a feature of the book, and I hope that the reader will agree

and feel free to skip around (see *How to use this book* below). I tried to "reboot" at the beginning of each part and again at the beginning of each chapter, to refresh our motivation. For researchers working with quaternion algebras, I have tried to collect results otherwise scattered in the literature and to provide some clarifications, corrections, and complete proofs in the hope that this text will provide a convenient reference. In order to provide these features, to the extent possible I have opted for an organizational pattern that is "horizontal" rather than "vertical": the text has many chapters, each representing a different slice of the theory.

I tried to compactify the text as much as possible, without sacrificing completeness. There were a few occasions when I thought a topic could use further elaboration or has evolved from the existing literature, but did not want to overburden the text; I collected these in a supplementary text *Quaternion algebras companion*, available at the website for the text at http://quatalg.org.

As usual, each chapter also contains a number of exercises at the end, ranging from checking basic facts used in a proof to more difficult problems that stretch the reader. Exercises that are used in the text are marked by ▹. For a subset of exercises (including many of those marked with ▹), there are hints, comments, or a complete solution available online.

How to Use This Book

With apologies to Whitman, this book is large, it contains multitudes—and hopefully, it does not contradict itself!

There is no obligation to read the book linearly cover to cover, and the reader is encouraged to find their own path, such as the following.

1. For an *introductory survey* course on quaternion algebras, read just the introductory sections in each chapter, those labeled with ▹, and supplement with sections from the text when interested. These introductions usually contain motivation and a summary of the results in the rest of the chapter, and I often restrict the level of generality or make simplifying hypotheses so that the main ideas are made plain. The reader who wants to quickly and gently grab hold of the basic concepts may digest the book in this way. The instructor may desire to fill in some further statements or proofs to make for a one semester course: Chapters 1, 2, 11, 25, and 35 could be fruitfully read in their entirety.
2. For a minicourse in *noncommutative algebra* with emphasis on quaternion algebras, read just Part I. Such an early graduate course would have minimal prerequisites and in a semester could be executed at a considered pace; it would provide the foundation for further study in many possible directions.
3. For quaternion algebras and *algebraic number theory*, read Part I and II. This course would be a nice second-semester addition following a standard first-semester course in algebraic number theory, suitable for graduate students in algebra and number theory who are motivated to study quaternion

algebras as "noncommutative quadratic fields". For a lighter course, Chapters 6, 20, and 21 could be skipped, and the instructor may opt to cover only the introductory section of a chapter for reasons of time and interest. To reinforce concepts from algebraic number theory, special emphasis could be placed on Chapter 13 (where local division algebras are treated like local fields) and 18 (where maximal orders are treated like noncommutative Dedekind domains).

There are also more specialized options, beginning with the introductory sections in Part I and continuing as follows:

4. For quaternion algebras and *analytic number theory*, continue with the introductory sections in Part II (just Chapters 9–17), and then cover Part III (at least through Chapter 29). This course could follow a first-semester course in analytic number theory, enriching students' understanding of zeta functions and L-functions (roughly speaking, beginning the move from GL_1 to GL_2). The additional prerequisite of real analysis (measure theory) is recommended. Optionally, this course could break after Chapter 26 to avoid adeles, and perhaps resume in an advanced topics course with the remaining chapters.
5. For quaternionic applications to *geometry* (specifically, hyperbolic geometry and low-dimensional topology), continue with the introductory sections in Part II (through Chapter 14), and then cover Part IV (optionally skipping Chapter 32).
6. For an advanced course on quaternion algebras and *arithmetic geometry*, continue with Part II, the introductory sections in Parts IV and V. Chapter 41 could be read immediately after Part II. This path is probably most appropriate for an advanced course for students with some familiarity with modular forms and some hyperbolic geometry, and Chapter 42 is probably only meaningful for students with a background in elliptic curves (though the relevant concepts are reviewed at the start).
7. Finally, for the reader who is studying quaternion algebras with an eye to applications with *supersingular elliptic curves*, the reader may follow Chapters 2–4, 9–10, 13–14, 16–17, 23, then the main event in Chapter 42. For further reading on quaternion orders and ternary quadratic forms, I suggest Chapters 5, 22, and 24.

Sections of the text that are more advanced (requiring more background) or those may be omitted are labeled with $*$. The final chapter (Chapter 43) is necessarily more advanced, and additional prerequisites in algebraic and arithmetic geometry are indicated.

It is a unique feature of quaternion algebras that topics overlap and fold together like this, and so I hope the reader will forgive the length of the book. The reader

may find the symbol definition list at the end to help in identifying unfamiliar notation. Finally, to ease in location, I have chosen to number all objects (theorem-like environments, equations, and figures) consecutively.

Companion Reading

Several general texts can serve as companion reading for this monograph:

- The lecture notes of Vignéras [Vig80a] have been an essential reference for the arithmetic of quaternion algebras since their publication.
- The seminal text by Reiner [Rei2003] on maximal orders treats many introductory topics that overlap this text.
- The book of Maclachlan–Reid [MR2003] gives an introduction to quaternion algebras with application to the geometry of 3-manifolds.
- The book by Deuring [Deu68] (in German) develops the theory of algebras over fields, culminating in the treatment of zeta functions of division algebras over the rationals, and may be of historical interest as well.
- Finally, Pizer [Piz76a] and Alsina–Bayer [AB2004] present arithmetic and algorithmic aspects of quaternion algebras over $\mathbb{Q}$.

Hanover, NH, USA John Voight

Acknowledgements

This book began as notes from a course offered at McGill University in the Winter 2010 semester, entitled *Computational aspects of quaternion algebras and Shimura curves*. I would like to thank the members of my Math 727 class for their invaluable discussions and corrections: Dylan Attwell-Duval, Xander Faber, Luis Finotti, Andrew Fiori, Cameron Franc, Adam Logan, Marc Masdeu, Jungbae Nam, Aurel Page, Jim Parks, Victoria de Quehen, Rishikesh, Shahab Shahabi, and Luiz Takei. This course was part of the special thematic semester *Number Theory as Applied and Experimental Science* organized by Henri Darmon, Eyal Goren, Andrew Granville, and Mike Rubinstein at the Centre de Recherche Mathématiques (CRM) in Montréal, Québec, and the extended visit was made possible by the generosity of Dominico Grasso, dean of the College of Engineering and Mathematical Sciences, and Jim Burgmeier, chair of the Department of Mathematics and Statistics, at the University of Vermont. I am very grateful for their support.

After a long hiatus, the writing continued while the author was on sabbatical at the University of California, Berkeley. Several students attended these lectures and gave helpful feedback: Watson Ladd, Andrew Niles, Shelly Manber, Eugenia Ro su, Emmanuel Tsukerman, Victoria Wood, and Alex Youcis. Thanks to Ken Ribet for sponsoring my visit and helping to organize these lectures. My sabbatical from Dartmouth College for the Fall 2013 and Winter 2014 quarters was made possible by the efforts of Associate Dean David Kotz, and I thank him for his support. During this sabbatical, further progress on the text was made in preparation for a minicourse on Brandt modules as part of *Minicourses on Algebraic and Explicit Methods in Number Theory*, organized by Cécile Armana and Christophe Delaunay at the Laboratoire de Mathématiques de Besançon in Salins-les-Bains, France.

Thanks also go to the patient participants in my Math 125 *Quaternion algebras* class at Dartmouth in Spring 2014 (Daryl Deford, Tim Dwyer, Zeb Engberg, Michael Firrisa, Jeff Hein, Nathan McNew, Jacob Richey, Tom Shemanske, Scott Smedinghoff, and David Webb) and my Math 125 *Geometry of discrete groups* in Summer 2015 (Angelica Babei, Ben Breen, Sara Chari, Melanie Dennis, James Drain, Tim Dwyer, Jon Epstein, David Freund, Sam Kater, Michael Musty, Nicolas

Petit, Sam Schiavone, Scott Smedinghoff, and Everett Sullivan). I am also grateful for the feedback given during my course at the Institute for Computational and Experimental Research in Mathematics (ICERM) at Brown University in Fall 2015, as part of the Semester Program on *Computational Aspects of the Langlands Program*: in particular, thanks to Henri Cohen, Edgar Costa, David Farmer, Winnie Li, David Roberts, and Drew Sutherland for their comments and questions.

During the preparation of this monograph, I was supported by the NSF Division of Mathematical Sciences Awards for *Quaternion algebras, Shimura curves, and modular forms: Algorithms and arithmetic* (DMS-0901971, July 2009–August 2011) and *CAREER: Explicit methods in arithmetic geometry* (DMS-1151047, July 2012–July 2018), and Simons Collaborations in Mathematics and the Physical Sciences Award, *Arithmetic geometry, number theory, and computation* (from September 2017). The Open Access publication of this book was made possible in part by generous support from Dartmouth College via Faculty Research and Professional Development funds, a John M. Manley Huntington Award, and Dartmouth Library Open Access Publishing.

It is somehow fitting that I would find myself writing this text while a faculty member at Dartmouth College: the story of the quaternions is interwoven with history of mathematics at Dartmouth. The only mathematical output by a Dartmouth professor in the nineteenth century was by Arthur Sherburne Hardy, Ph. D., the author of an 1881 text on quaternions entitled *Elements of quaternions* [Har1881]. Brown describes it as

> an adequate, if not inspiring text. It was something for Dartmouth to offer a course in such an abstruse field, and the course was actually given a few times when a student and an instructor could be found simultaneously [Bro61, p. 2].

I can only hope that this book will receive better reviews!

On a more personal note, I have benefited from the insight of incredible teachers and collaborators throughout my mathematical life, and their positive impact can hopefully be seen in this manuscript: among many, I would like to give special thanks to Pete L. Clark, Bas Edixhoven, Benedict Gross, Hendrik Lenstra, and Bjorn Poonen for their inspiration and guidance over the years.

Many thanks go to the others who offered helpful comments, corrections, answered my questions, and provided feedback: Konstantin Ardakov, Sarah Arpin, Asher Auel, Andreas Bächle, John Boxall, Ben Breen, Juliusz Brzezinski, Billy Chan, Pierre Clare, France Dacar, Lassina Dembélé, Juanita Duque Rosero, Kirsten Eisentraeger, John Enns, Zak Evans, Steven Galbraith, Darij Grinberg, Joseph Gunther, Jonathan Hanke, Thorsten Herrig, Aleksander Horawa, Will Jagy, Ariyan Javanpeykar, BoGwang Jeon, Chan-Ho Kim, Robert Kucharczyk, Joel Laity, Dion Leijnse, Patrick Lenning, Chao Li, Wen-Wei Li, Benjamin Linowitz, Travis Morrison, Ariel Pacetti, Lorenz Panny, Jennifer Park, Carl Pomerance, Jose Pujol, Matthieu Rambaud, John Riccardi, Sam Roven, Nithi Rungtanapirom, Ciaran Schembri, sibilant, Jana Sotakova, Tom Shemanske, Daniel Smertnig, Jim Stankewicz, Kate Stange, Drew Sutherland, Nicole Sutherland, Jacob Swenberg, Anne van Weerden, Annalena Wernz, Mckenzie West, Dana Williams, Jiangwei

Xue, and Vitthal Yellambalse. I am indebted to C. Emily I. Redelmeier for fabulous Technical assistance with figures. Special thanks to the anonymous referees for their constructive feedback and to Sara Chari, Peter Landweber, Joe Quinn, and Grant Molnar for a careful reading. The errors and omissions that remain are, of course, my own: please contact me at jvoight@gmail.com if you find mistakes or have other suggestions (e.g., additional references, alternate proofs, further exercises or applications). Collected errata and a corrected version will be posted on the webpage for the text (http://quatalg.org).

I am profoundly grateful to those who lent encouragement at various times during the writing of this book, when the going was tough: Srinath Baba, Chantal David, Matthew Greenberg and Kristina Loeschner, Laurie Johnson at the National Center for Faculty Development and Diversity (NCFDD), David Michaels, Tom Shemanske, and my mother Connie Voight. Thank you all! I am also grateful to Marc Strauss and Loretta Bartolini, editors at Springer, for their professional guidance. Finally, I would like to offer my deepest gratitude to my partner Brian Kennedy: this book would not have been possible without his patience, equanimity, and wit. As they say, happiness is a journey, not a destination; the same I think is true for happy mathematics, and I am blessed to have had both personal and professional companionship on this one.

Contents

II Arithmetic

Chapter 1

Introduction

We begin following the historical arc of quaternion algebras and tracing their impact on the development of mathematics. Our account is selective: for further overview, see Lam [Lam2003] and Lewis [Lew2006a].

1.1 Hamilton's quaternions

In perhaps the "most famous act of mathematical vandalism", on October 16, 1843, Sir William Rowan Hamilton (1805–1865, Figure 1.1.2) carved the following equations into the Brougham Bridge (now Broom Bridge) in Dublin:

$$i^2 = j^2 = k^2 = ijk = -1. \tag{1.1.1}$$

His discovery was a defining moment in the history of algebra (Figure 1.1.3).

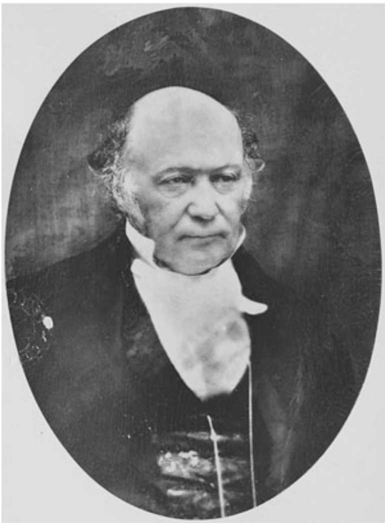


Figure 1.1.2: William Rowan Hamilton
(public domain; scan by Wellesley College Library)

© The Author(s) 2021

J. Voight, *Quaternion Algebras*, Graduate Texts in Mathematics 288,
https://doi.org/10.1007/978-3-030-56694-4_1

For at least ten years (on and off), Hamilton had been attempting to model (real) three-dimensional space with a structure like the complex numbers, whose addition and multiplication occur in two-dimensional space. Just like the complex numbers had a "real" and "imaginary" part, so too did Hamilton hope to find an algebraic system whose elements had a "real" and two-dimensional "imaginary" part. In the early part of the month of October 1843, his sons Archibald Henry and William Edwin Hamilton, while still very young, would ask their father at breakfast [Ham67, p. xv]: "Well, papa, can you multiply triplets?" To which Hamilton would reply, "with a sad shake of the head, 'No, I can only add and subtract them'" [Ham67, p. xv]. For a history of the "multiplying triplets" problem—the nonexistence of division algebra over the reals of dimension 3—see May [May66, p. 290].



Figure 1.1.3: *William Rowan Hamilton*, a sand sculpture by Daniel Doyle, part of the 2012 Dublin castle exhibition, *Irish Science* (© Daniel Doyle; reproduced with permission)

Then, on the dramatic day in 1843, Hamilton's had a flash of insight [Ham67, p. xv–xvi], which he described in a letter to Archibald (written in 1865):

> On the 16th day of [October]—which happened to be a Monday, and a Council day of the Royal Irish Academy—I was walking in to attend and preside, and your mother was walking with me, along the Royal Canal, to which she had perhaps driven; and although she talked with me now and then, yet an under-current of thought was going on in my mind, which gave at last a result, whereof it is not too much to say that I felt at once the importance. An electric circuit seemed to close; and a spark flashed forth, the herald (as I foresaw, immediately) of many long years to come of definitely directed thought and work, by myself if spared, and at all

> events on the part of others, if I should even be allowed to live long enough distinctly to communicate the discovery. Nor could I resist the impulse—unphilosophical as it may have been—to cut with a knife on a stone of Brougham Bridge, as we passed it, the fundamental formula with the symbols, i, j, k; namely,
>
> $$i^2 = j^2 = k^2 = ijk = -1$$
>
> which contains the Solution of the Problem, but of course, as an inscription, has long since mouldered away.

In this moment, Hamilton realized that he needed a fourth dimension; he later coined the term *quaternions* for the real space spanned by the elements $1, i, j, k$, subject to his multiplication laws. He presented his theory of quaternions to the Royal Irish Academy in a paper entitled "On a new Species of Imaginary Quantities connected with a theory of Quaternions" [Ham1843]. Today, we denote this algebra $\mathbb{H} := \mathbb{R} + \mathbb{R}i + \mathbb{R}j + \mathbb{R}k$ and call $\mathbb{H}$ the ring of *Hamilton quaternions* in his honor.

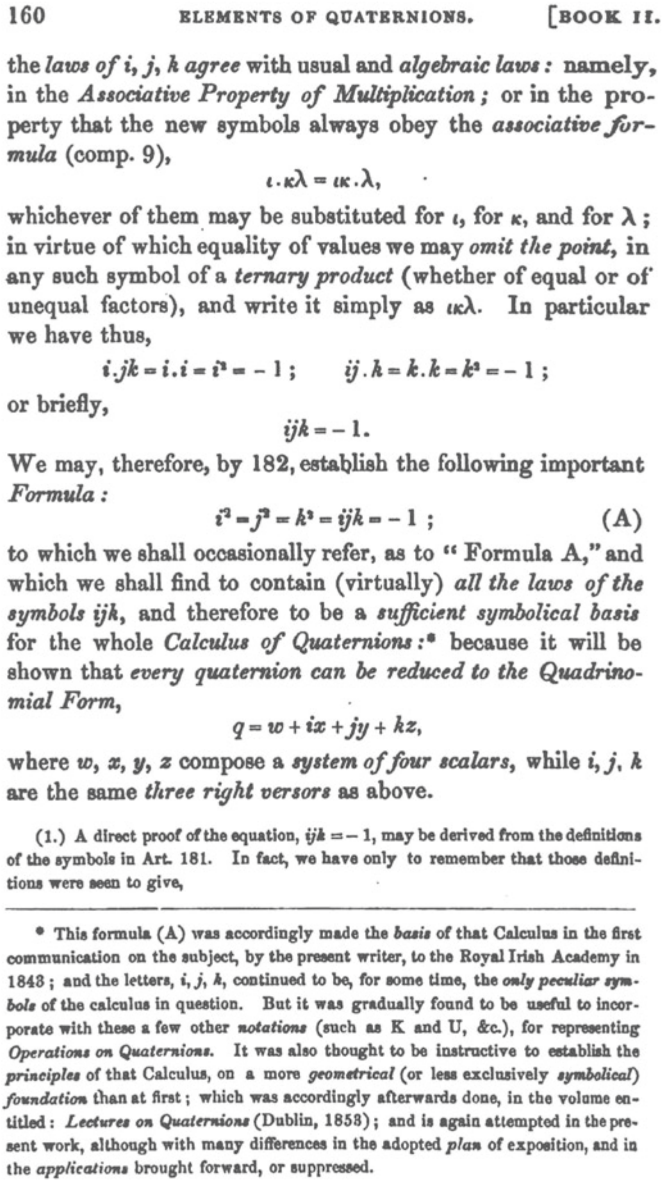
This charming story of quaternionic discovery remains in the popular consciousness, and to commemorate Hamilton's discovery of the quaternions, there is an annual "Hamilton walk" in Dublin [ÓCa2010]. Although his carvings have long since worn away, a plaque on the bridge now commemorates this significant event in mathematical history (Figure 1.1.4).



Figure 1.1.4: The Broom Bridge plaque (author's photo)

For more on the history of Hamilton's discovery, see the extensive and detailed accounts of Dickson [Dic19] and Van der Waerden [vdW76]. There are also three main biographies written about the life of William Rowan Hamilton, a man sometimes referred to as "Ireland's greatest mathematician": by Graves [Grav1882, Grav1885, Grav1889] in three volumes, Hankins [Hankin80], and O'Donnell [O'Do83]. Numer-

ous other shorter biographies have been written [DM89, Lanc67, ÓCa2000]. (Certain aspects of Hamilton's private life deserve a more positive portrayal, however: see Van Weerden–Wepster [WW2018].)



160 ELEMENTS OF QUATERNIONS. [BOOK II.

the *laws of i, j, k agree* with usual and *algebraic laws:* namely, in the *Associative Property of Multiplication;* or in the property that the new symbols always obey the *associative formula* (comp. 9),

$$\iota . \kappa\lambda = \iota\kappa . \lambda,$$

whichever of them may be substituted for ι, for κ, and for λ; in virtue of which equality of values we may *omit the point,* in any such symbol of a *ternary product* (whether of equal or of unequal factors), and write it simply as $\iota\kappa\lambda$. In particular we have thus,

$$i.jk = i.i = i^2 = -1; \qquad ij.k = k.k = k^2 = -1;$$

or briefly,

$$ijk = -1.$$

We may, therefore, by 182, establish the following important *Formula:*

$$i^2 = j^2 = k^2 = ijk = -1; \qquad \text{(A)}$$

to which we shall occasionally refer, as to "Formula A," and which we shall find to contain (virtually) *all the laws of the symbols ijk,* and therefore to be a *sufficient symbolical basis* for the whole *Calculus of Quaternions:** because it will be shown that *every quaternion can be reduced to the Quadrinomial Form,*

$$q = w + ix + jy + kz,$$

where w, x, y, z compose a *system of four scalars,* while i, j, k are the same *three right versors* as above.

(1.) A direct proof of the equation, $ijk = -1$, may be derived from the definitions of the symbols in Art. 181. In fact, we have only to remember that those definitions were seen to give,

* This formula (A) was accordingly made the *basis* of that Calculus in the first communication on the subject, by the present writer, to the Royal Irish Academy in 1843; and the letters, i, j, k, continued to be, for some time, the *only peculiar symbols* of the calculus in question. But it was gradually found to be useful to incorporate with these a few other *notations* (such as K and U, &c.), for representing *Operations on Quaternions.* It was also thought to be instructive to establish the *principles* of that Calculus, on a more *geometrical* (or less exclusively *symbolical*) *foundation* than at first; which was accordingly afterwards done, in the volume entitled: *Lectures on Quaternions* (Dublin, 1853); and is again attempted in the present work, although with many differences in the adopted *plan* of exposition, and in the *applications* brought forward, or suppressed.

Figure 1.1.5: A page from Hamilton's *Elements of quaternions* [Ham1866] (public domain)

There are several precursors to Hamilton's discovery that bear mentioning. First, the quaternion multiplication laws are already implicit in the four-square identity of

Leonhard Euler (1707–1783):

$$\begin{aligned}(a_1^2 + a_2^2 + a_3^2 + a_4^2)(b_1^2 + b_2^2 + b_3^2 + b_4^2) = c_1^2 + c_2^2 + c_3^2 + c_4^2 = \\ (a_1b_1 - a_2b_2 - a_3b_3 - a_4b_4)^2 + (a_1b_2 + a_2b_1 + a_3b_4 - a_4b_3)^2 \\ + (a_1b_3 - a_2b_4 + a_3b_1 + a_4b_2)^2 + (a_1b_4 + a_2b_3 - a_3b_2 + a_4b_1)^2.\end{aligned} \tag{1.1.6}$$

Indeed, the full multiplication law for quaternions reads precisely

$$(a_1 + a_2i + a_3j + a_4k)(b_1 + b_2i + b_3j + b_4k) = c_1 + c_2i + c_3j + c_4k$$

with c_1, c_2, c_3, c_4 as defined in (1.1.6); the four-square identity corresponds to taking a norm on both sides.

It was perhaps Carl Friedrich Gauss (1777–1855) who first observed this connection. In a note dated around 1819 [Gau00], he interpreted the formula (1.1.6) as a way of composing real quadruples: to the quadruples (a_1, a_2, a_3, a_4) and (b_1, b_2, b_3, b_4) in $\mathbb{R}^4$, he defined the composite tuple (c_1, c_2, c_3, c_4) and noted the noncommutativity of this operation. Gauss elected not to publish these findings (as he chose not to do with many of his discoveries). In letters to De Morgan [Grav1885, Grav1889, p. 330, p. 490], Hamilton attacks the allegation that Gauss had discovered quaternions first.

Finally, Olinde Rodrigues (1795–1851) (of the *Rodrigues formula* for Legendre polynomials) gave a formula for the angle and axis of a rotation in $\mathbb{R}^3$ obtained from two successive rotations—essentially giving a different parametrization of the quaternions—but had left mathematics for banking long before the publication of his paper [Rod1840]. The story of Rodrigues and the quaternions is given by Altmann [Alt89] and Pujol [Puj2012], and the fuller story of his life is recounted by Altmann–Ortiz [AO2005]. See also the description by Pujol [Puj2014] of Hamilton's derivation of the relation between rotations and quaternions from 1847, set in historical context.

In any case, the quaternions consumed the rest of Hamilton's academic life and resulted in the publication of two bulky treatises [Ham1853, Ham1866] (see also the review [Ham1899]). Hamilton's mathematical writing over these years, an example of which can be found in Figure 1.1.5, was at times opaque; nevertheless, many physicists used quaternions extensively and for a long time in the mid-19th century, quaternions were an essential notion in physics.

Other figures contemporaneous with Hamilton were also developing vectorial systems, most notably Hermann Grassmann (1809–1877) [Gras1862]. The modern notion of vectors was developed by Willard Gibbs (1839–1903) and Oliver Heaviside (1850–1925), independently. In 1881 and 1884 (in two halves), Gibbs introduced in a pamphlet *Elements of Vector Analysis* the now standard vector notation of the cross product and dot product, with the splendid equality

$$vw = -v \cdot w + v \times w \tag{1.1.7}$$

for $v, w \in \mathbb{R}i + \mathbb{R}j + \mathbb{R}k \subset \mathbb{H}$ relating quaternionic multiplication on the left to dot and cross products on the right. (The equality (1.1.7) also appears in Hamilton's work, but in different notation.) Gibbs did not consider the quaternion product to be a "*fundamental* notion in vector analysis" [Gib1891, p. 512], and argued for a vector analysis that

would apply in arbitrary dimension; on the relationship between these works, Gibbs wrote after learning of the work of Grassmann: "I saw that the methods wh[ich] I was using, while nearly those of Hamilton, were almost exactly those of Grassmann" [Whe62, p. 108]. For more on the history of quaternionic and vector calculus, see Crowe [Cro64] and Simons [Sim2010].

The rivalry between physical notations flared into a war in the latter part of the 19th century between the 'quaternionists' and the 'vectorists', and for some the preference of one system versus the other became an almost partisan split. On the side of quaternions, James Clerk Maxwell (1831–1879), who derived the equations which describe electromagnetic fields, wrote [Max1869, p. 226]:

> The invention of the calculus of quaternions is a step towards the knowledge of quantities related to space which can only be compared, for its importance, with the invention of triple coordinates by Descartes. The ideas of this calculus, as distinguished from its operations and symbols, are fitted to be of the greatest use in all parts of science.

And Peter Tait (1831–1901), Hamilton's "chief disciple" [Hankin80, p. 316], wrote in 1890 [Tai1890] decrying notation and attacking Willard Gibbs (1839–1903):

> It is disappointing to find how little progress has recently been made with the development of Quaternions. One cause, which has been specially active in France, is that workers at the subject have been more intent on modifying the notation, or the mode of presentation of the fundamental principles, than on extending the applications of the Calculus. . . . Even Prof. Willard Gibbs must be ranked as one the retarders of quaternions progress, in virtue of his pamphlet on *Vector Analysis*, a sort of hermaphrodite monster, compounded of the notation of Hamilton and Grassman.

Game on! On the vectorist side, Lord Kelvin (a.k.a. William Thomson, who formulated the laws of thermodynamics), said in an 1892 letter to R. B. Hayward about his textbook in algebra (quoted in Thompson [Tho10, p. 1070]):

> Quaternions came from Hamilton after his really good work had been done; and, though beautifully ingenious, have been an unmixed evil to those who have touched them in any way, including Clerk Maxwell.

(There is also a rompous fictionalized account by Pynchon in his tome *Against the Day* [Pyn2006].) Ultimately, the superiority and generality of vector notation carried the day, and only certain useful fragments of Hamilton's quaternionic notation—e.g., the "right-hand rule" $i \times j = k$ in multivariable calculus—remain in modern usage.

1.2 Algebra after the quaternions

The debut of Hamilton's quaternions was met with some resistance in the mathematical world: it proposed a system of "numbers" that did not satisfy the usual commutative rule of multiplication. Quaternions predated even the notion of matrices, introduced in 1855 by Arthur Cayley (1821–1895). Hamilton's bold proposal of a noncommutative

multiplication law was the harbinger of a burgeoning array of algebraic structures. In the words of J.J. Sylvester [Syl1883, pp. 271–272]:

> In Quaternions (which, as will presently be seen, are but the simplest order of matrices viewed under a particular aspect) the example had been given of Algebra released from the yoke of the commutative principle of multiplication—an emancipation somewhat akin to Lobachevsky's of Geometry from Euclid's noted empirical axiom; and later on, the Peirces, father and son (but subsequently to 1858) had prefigured the universalization of Hamilton's theory, and had emitted an opinion to the effect that probably all systems of algebraical symbols subject to the associative law of multiplication would be eventually found to be identical with linear transformations of schemata susceptible of matriculate representation.

So with the introduction of the quaternions, the floodgates of algebraic possibility had been opened. See Happel [Hap80] for an overview of the early development of algebra following Hamilton's quaternions, as well as the more general history given by Van der Waerden [vdW85, Chapters 10–11].

The day after his discovery, Hamilton sent a letter [Ham1844] describing the quaternions to his friend John T. Graves (1806–1870). Graves replied on October 26, 1843, with his compliments, but added:

> There is still something in the system which gravels me. I have not yet any clear views as to the extent to which we are at liberty arbitrarily to create imaginaries, and to endow them with supernatural properties. . . . If with your alchemy you can make three pounds of gold, why should you stop there?

Following through on this invitation, on December 26, 1843, Graves wrote to Hamilton that he had successfully generalized the quaternions to the "octaves", now called *octonions* $\mathbb{O}$, an algebra in eight dimensions, with which he was able to prove that the product of two sums of eight perfect squares is another sum of eight perfect squares, a formula generalizing (1.1.6). In fact, Hamilton first invented the term *associative* in 1844, around the time of his correspondence with Graves. Unfortunately for Graves, the octonions were discovered independently and published in 1845 by Cayley [Cay1845b], who often is credited for their discovery. (Even worse, the eight squares identity was also previously discovered by C. F. Degen.) For a more complete account of this story and the relationships between quaternions and octonions, see the survey article by Baez [Bae2002], the article by Van der Blij [vdB60], and the delightful book by Conway–Smith [CSm2003].

Cayley also studied quaternions themselves [Cay1845a] and was able to reinterpret them as arising from a *doubling process*, also called the *Cayley–Dickson construction*, which starting from $\mathbb{R}$ produces $\mathbb{C}$ then $\mathbb{H}$ then $\mathbb{O}$, taking the ordered, commutative, associative algebra $\mathbb{R}$ and progressively deleting one adjective at a time. So algebras were first studied over the real and complex numbers and were accordingly called *hypercomplex numbers* in the late 19th and early 20th century. And this theory flourished. Hamilton himself considered the algebra over $\mathbb{C}$ defined by his famous

equations (1.1.1), calling them *biquaternions*. In 1878, Ferdinand Frobenius (1849–1917) proved that the only finite-dimensional associative real division algebras are $\mathbb{R}$, $\mathbb{C}$, and $\mathbb{H}$ [Fro1878]. This result was also proven independently by C.S. Peirce, the son of Benjamin Peirce, below. Adolf Hurwitz (1859–1919) later showed that the only *normed* finite-dimensional not-necessarily-associative real division algebras are $\mathbb{R}$, $\mathbb{C}$, $\mathbb{H}$, and $\mathbb{O}$. (The same statement is true without the condition that the algebra be normed, but currently the proofs use topology, not algebra! Bott–Milnor [BM58] and Kervaire [Ker58] proved that the $(n-1)$-dimensional sphere $\{x \in \mathbb{R}^n : \|x\|^2 = 1\}$ has trivial tangent bundle if and only if there is an n-dimensional not-necessarily-associative real division algebra if and only if $n = 1, 2, 4, 8$. The solution to the Hopf invariant one problem by Adams also implies this result; an elegant and concise proof using K-theory, Adams operations, and elementary number theory was given by Adams–Atiyah [AA66]. See Hirzebruch [Hir91] or Ranicki [Ran2011] for a more complete account.)

In another attempt to seek a generalization of the quaternions to higher dimension, William Clifford (1845–1879) developed a way to build algebras from quadratic forms in 1876 [Cli1878]. Clifford constructed what we now call a *Clifford algebra* $C(V)$ associated to $V = \mathbb{R}^n$ (with the standard Euclidean norm); it is an algebra of dimension 2^n containing V with multiplication induced from the relation $x^2 = -\|x\|^2$ for all $x \in V$. We have $C(\mathbb{R}^1) = \mathbb{C}$ and $C(\mathbb{R}^2) = \mathbb{H}$, so the Hamilton quaternions arise as a Clifford algebra—but $C(\mathbb{R}^3)$ is not the octonions. The theory of Clifford algebras tightly connects the theory of quadratic forms and the theory of normed division algebras and its impact extends in many mathematical directions. For more on the history of Clifford algebras, see Diek–Kantowski [DK95].

A further physically motivated generalization was pursued by Alexander Macfarlane (1851–1913): he developed a theory of what he called *hyperbolic quaternions* [Macf00] (a revised version of an earlier, nonassociative attempt [Macf1891]), with the multiplication laws

$$\begin{gathered} i^2 = j^2 = k^2 = 1, \\ ij = \sqrt{-1}k = -ji, \quad jk = \sqrt{-1}i = -kj, \quad ki = \sqrt{-1}j = -ik. \end{gathered} \tag{1.2.1}$$

Thought of as an algebra over $\mathbb{C} = \mathbb{R}(\sqrt{-1})$, Macfarlane's hyperbolic quaternions are isomorphic to Hamilton's biquaternions (and therefore isomorphic to $\mathrm{M}_2(\mathbb{C})$). Moreover, the restriction of the norm to the real span of the basis $1, i, j, k$ in Macfarlane's algebra is a quadratic form of signature $(1, 3)$: this gives a quaternionic version of space-time, something also known as *Minkowski space* (but with Macfarlane's construction predating that of Minkowski). For more on the history and further connections, see Crowe [Cro64].

Around this time, other types of algebras over the real numbers were also being investigated, the most significant of which were Lie algebras. In the seminal work of Sophus Lie (1842–1899), group actions on manifolds were understood by looking at this action infinitesimally; one thereby obtains a *Lie algebra* of vector fields that determines the local group action. The simplest nontrivial example of a Lie algebra is the cross product of two vectors, related to quaternion multiplication in (1.1.7): it defines, a linear, alternating, but nonassociative binary operation on $\mathbb{R}^3$ that satisfies

the Jacobi identity emblematized by

$$i \times (j \times k) + k \times (i \times j) + j \times (k \times i) = 0. \tag{1.2.2}$$

The Lie algebra "linearizes" the group action and is therefore more accessible. Wilhelm Killing (1847–1923) initiated the study of the classification of Lie algebras in a series of papers [Kil1888], and this work was completed by Élie Cartan (1869–1951). We refer to Hawkins [Haw2000] for a description of this rich series of developments.

In this way, the study of division algebras gradually evolved, independent of physical interpretations. Benjamin Peirce (1809–1880) in 1870 developed what he called *linear associative algebras* [Pei1882]; he provided a decomposition of an algebra relative to an *idempotent* (his terminology). The first definition of an algebra over an arbitrary field seems to have been given by Leonard E. Dickson (1874–1954) [Dic03]: at first he still called the resulting object a *system of complex numbers* and only later adopted the name *linear algebra*.

The notion of a *simple* algebra had been discovered by Cartan, and Theodor Molien (1861–1941) had earlier shown in his terminology that every simple algebra over the complex numbers is a matrix algebra [Mol1893]. But it was Joseph Henry Maclagan Wedderburn (1882–1948) who was the first to find meaning in the structure of simple algebras over an arbitrary field, in many ways leading the way forward. The jewel of his 1908 paper [Wed08] is still foundational in the structure theory of algebras: a simple algebra (finite-dimensional over a field) is isomorphic to a matrix ring over a division ring. Wedderburn also proved that a finite division ring is a field, a result that like his structure theorem has inspired much mathematics. For more on the legacy of Wedderburn, see Artin [Art50].

In the early 1900s, Dickson was the first to consider quaternion algebras over a general field [Dic12, (8), p. 65]. He began by considering more generally those algebras in which every element satisfies a quadratic equation [Dic12], exhibited a diagonalized basis for such an algebra, and considered when such an algebra can be a division algebra. This led him to multiplication laws for what he later called a *generalized quaternion algebra* [Dic14, Dic23], with multiplication laws

$$\begin{gathered} i^2 = a, \quad j^2 = b, \quad k^2 = -ab, \\ ij = k = -ji, \quad ik = aj = -ki, \quad kj = bi = -jk \end{gathered} \tag{1.2.3}$$

with a, b nonzero elements in the base field. (To keep track of these, it is helpful to write i, j, k around a circle clockwise.) Today, we no longer employ the adjective "generalized"—over fields other than $\mathbb{R}$, there is no reason to privilege the Hamiltonian quaternions—and we can reinterpret this vein of Dickson's work as showing that every 4-dimensional central simple algebra is a quaternion algebra (a statement that holds even over a field F with $\operatorname{char} F = 2$). See Fenster [Fen98] for a summary of Dickson's work in algebra, and Lewis [Lew2006b] for a broad survey of the role of involutions and anti-automorphisms in the classification of algebras.

1.3 Quadratic forms and arithmetic

Hamilton's quaternions also fused a link between quadratic forms and arithmetic, phrased in the language of noncommutative algebra. Indeed, part of Dickson's interest in quaternion algebras stemmed from earlier work of Hurwitz [Hur1898], alluded to above. Hurwitz had asked for generalizations of the composition laws arising from sum of squares laws like that of Euler (1.1.6) for four squares and Cayley for eight squares: for which n does there exist an identity

$$(a_1^2 + \cdots + a_n^2)(b_1^2 + \cdots + b_n^2) = c_1^2 + \cdots + c_n^2$$

with each c_i bilinear in the variables a and b? He then proved [Hur1898] that over a field where 2 is invertible, these identities exist only for $n = 1, 2, 4, 8$ variables (so in particular, there is no formula expressing the product of two sums of 16 squares as the sum of 16 squares). As Dickson [Dic19] further explained, this result of Hurwitz is intimately tied to the theory of algebras. For more on compositions of quadratic forms and their history, including theorems of Hurwitz–Radon and Pfister, see Shapiro [Sha90].

Thinking along similar lines, Hurwitz gave a new proof of the four-square theorem of Lagrange, that every positive integer is the sum of four integer squares: he first wrote about this in 1896 on quaternionic number theory ("Über die Zahlentheorie der Quaternionen") [Hur1896], then published a short book on the subject in 1919 [Hur19]. To this end, Hurwitz considered Hamilton's equations over the rational numbers and said that a quaternion $t+xi+yj+zk$ with $t, x, y, z \in \mathbb{Q}$ was an *integer* if t, x, y, z all belonged to $\mathbb{Z}$ or all to $\frac{1}{2}+\mathbb{Z}$, conditions for the quaternion to satisfy a quadratic polynomial with integer coefficients. Hurwitz showed that his ring of integer quaternions, today called the *Hurwitz order*, admits a generalization of the Euclidean algorithm and thereby a factorization theory. He then applied this to count the number of ways of representing an integer as the sum of four squares, a result due to Jacobi. The notion of integral quaternions was also explored in the 1920s by Venkov [Ven22, Ven29] and the 1930s by Albert [Alb34]. Dickson considered further questions of representing positive integers by integral quaternary quadratic forms [Dic19, Dic23, Dic24] in the same vein.

So by the end of the 1920s, quaternion algebras were used to study quadratic forms in a kind of noncommutative algebraic number theory [Lat26, Gri28]. It was known that a (generalized) quaternion algebra (1.2.3) was semisimple in the sense of Wedderburn, and thus it was either a division algebra or a full matrix algebra over the ground field. Indeed, a quaternion algebra is a matrix algebra if and only if a certain ternary quadratic form has a nontrivial zero, and over the rational numbers this problem was already studied by Legendre. Helmut Hasse (1898–1979) reformulated Legendre's conditions: a quadratic form has a nontrivial zero over the rationals if and only if it has a nontrivial zero over the real numbers and Hensel's field of p-adic numbers for all odd primes p. This result paved the way for many further advances, and it is now known as *the Hasse principle* or the *local-global principle* for quadratic forms. For an overview of this history, see Scharlau [Scha2009, §1].

Further deep results in number theory were soon to follow. Dickson [Dic14] had defined *cyclic algebras*, reflecting many properties of quaternion algebras, and in

1929 lectures Emmy Noether (1882–1935) considered the even more general *crossed product algebras*. Not very long after, in a volume dedicated to Hensel's seventieth birthday, Richard Brauer (1901–1977), Hasse, and Noether proved a fundamental theorem for the structure theory of algebras over number fields [BHN31]: every central division algebra over a number field is a cyclic algebra. This crucial statement had profound implications for *class field theory*, the classification of abelian extensions of a number field, with a central role played by the *Brauer group* of a number field, a group encoding its division algebras. For a detailed history and discussion of these lines, see Fenster–Schwermer [FS2007], Roquette [Roq2006], and the history of class field theory summarized by Hasse himself [Hass67].

At the same time, Abraham Adrian Albert (1905–1972), a doctoral student of Dickson, was working on the structure of division algebras and algebras with involution, and he had written a full book on the subject [Alb39] collecting his work in the area, published in 1939. Albert had examined the tensor product of two quaternion algebras, called a *biquaternion algebra* (not to be confused with Hamilton's biquaternions), and he characterized when such an algebra was a division algebra in terms of a senary (six variable) quadratic form. Albert's classification of algebras with involution was motivated by understanding possible endomorphism algebras of abelian varieties, viewed as multiplier rings of Riemann matrices and equipped with the *Rosati involution*: a consequence of this classification is that quaternion algebras are the only noncommutative endomorphism algebras of simple abelian varieties. He also proved that a central simple algebra admits an involution if and only if the algebra is isomorphic to its opposite algebra (equivalently, it has order at most 2 in the Brauer group). For a biography of Albert and a survey of his work, see Jacobson [Jacn74]. Roquette argues convincingly [Roq2006, §8] that because of Albert's contributions to its proof (for example, his work with Hasse [AH32]), we should refer to the *Albert–Brauer–Hasse–Noether theorem* in the previous paragraph.

1.4 Modular forms and geometry

Quaternion algebras also played a formative role in what began as a subfield of complex analysis and ordinary differential equations and then branched into the theory of modular forms—and ultimately became a central area of modern number theory.

Returning to a thread from the previous section, the subject of representing numbers as the sum of four squares saw considerable interest in the 17th and 18th centuries [Dic71, Chapter VIII]. Carl Jacobi (1804–1851) approached the subject from the analytic point of view of theta functions, the basic building blocks for elliptic functions; these were first studied in connection with the problem of the arc length of an ellipse, going back to Abel. Jacobi studied the series

$$\theta(\tau) := \sum_{n=-\infty}^{\infty} \exp(2\pi i n^2 \tau) = 1 + 2q + 2q^4 + 2q^9 + \dots \tag{1.4.1}$$

where τ is a complex number with positive imaginary part and $q = \exp(2\pi i \tau)$. Jacobi

proved the remarkable identity

$$\theta(\tau)^4 = \sum_{a,b,c,d\in\mathbb{Z}} q^{a^2+b^2+c^2+d^2} = 1 + 8\sum_{n=1}^{\infty} \sigma^*(n)q^n, \tag{1.4.2}$$

where $\sigma^*(n) := \sum_{4\nmid d\mid n} d$ is the sum of divisors of n not divisible by 4. In this way, Jacobi gave an explicit formula for the number of ways of expressing a number as the sum of four squares. For a bit of history and an elementary derivation in the style of Gauss and Jacobi, see Ewell [Ewe82].

As a Fourier series, the Jacobi theta function θ (1.4.1) visibly satisfies $\theta(\tau + 1) = \theta(\tau)$. Moreover, owing to its symmetric description, Jacobi showed using Poisson summation that θ also satisfies the transformation formula

$$\theta(-1/\tau) = \sqrt{\tau/i}\,\theta(\tau). \tag{1.4.3}$$

Felix Klein (1849–1925) saw geometry in formulas like (1.4.3). In his *Erlangen Program* (1872), he recast 19th century geometry in terms of the underlying group of symmetries, unifying Euclidean and non-Euclidean formulations. Turning then to hyperbolic geometry, he studied the *modular group* $\mathrm{SL}_2(\mathbb{Z})$ acting by linear fractional transformations on the upper half-plane, and interpreted transformation formulas for elliptic functions: in particular, Klein defined his *absolute invariant* $J(\tau)$ [Kle1878], a function invariant under the modular group. Together with his student Robert Fricke (1861–1930), this led to four volumes [FK1890-2, FK1897, FK12] on elliptic modular functions and automorphic functions, combining brilliant advances in group theory, number theory, geometry, and invariant theory (Figure 1.4.4).

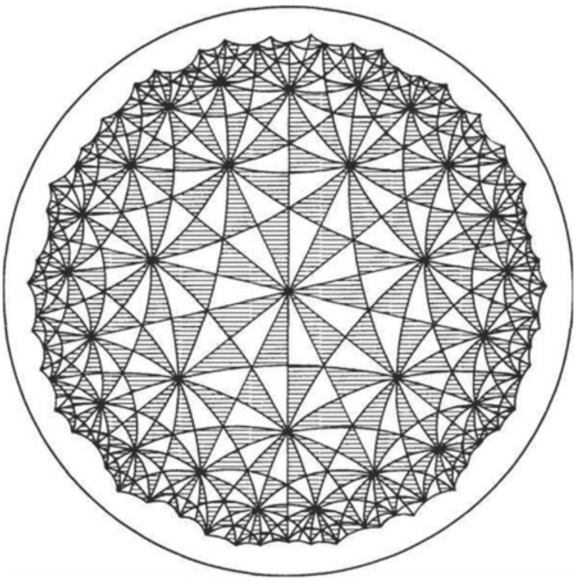


Figure 1.4.4: The (2, 3, 7)-tiling by Fricke and Klein [FK1890-2]
(public domain)

At the same time, Henri Poincaré (1854–1912) brought in the theory of linear differential equations—and a different, group-theoretic approach. In correspondence with Fuchs in 1880 on hypergeometric differential equations, he writes about the beginnings of his discovery of a new class of analytic functions [Gray2000, p.177]:

> They present the greatest analogy with elliptic functions, and can be represented as the quotient of two infinite series in infinitely many ways. Amongst those series are those which are entire series playing the role of Theta functions. These converge in a certain circle and do not exist outside it, as thus does the Fuchsian function itself. Besides these functions there are others which play the same role as the zeta functions in the theory of elliptic functions, and by means of which I solve linear differential equations of arbitrary orders with rational coefficients whenever there are only two finite singular points and the roots of the three determinantal equations are commensurable.

As he reminisced later in his *Science et Méthode* [Poi1908, p. 53]:

> I then undertook to study some arithmetical questions without any great result appearing and without expecting that this could have the least connection with my previous researches. Disgusted with my lack of success, I went to spend some days at the sea-side and thought of quite different things. One day, walking along the cliff, the idea came to me, always with the same characteristics of brevity, suddenness, and immediate certainty, that the arithmetical transformations of ternary indefinite quadratic forms were identical with those of non-Euclidean geometry.

In other words, like Klein, Poincaré launched a program to study complex analytic functions defined on the unit disc that are invariant with respect to a discrete group of matrix transformations that preserve a rational indefinite ternary quadratic form. Today, such groups are called *arithmetic Fuchsian groups*, and we study them as unit groups of quaternion algebras. To read more on the history of differential equations in the time of Riemann and Poincaré, see the history by Gray [Gray2000], as well as Gray's scientific biography of Poincaré [Gray2013].

In the context of these profound analytic discoveries, Erich Hecke (1887–1947) began his study of modular forms. He studied the Dedekind zeta function, a generalization of Riemann's zeta function to number fields, and established its functional equation using theta functions. In the study of similarly defined analytic functions arising from modular forms, he was led to define the "averaging" operators acting on spaces of modular forms that now bear his name. In this way, he could interpret the Fourier coefficients $a(n)$ of a Hecke eigenform (normalized, weight 2) as eigenvalues of his operators: he proved that they satisfy a relation of the form

$$a(m)a(n) = \sum_{d \mid \gcd(m,n)} a(mn/d^2)d \tag{1.4.5}$$

and consequently a two-term recursion relation. He thereby showed that the Dirichlet L-series of an eigenform, defined via Mellin transform, has an Euler product, analytic continuation, and functional equation.

Hecke went further, and connected the analytic theory of modular forms and his operators to the arithmetic theory of quadratic forms. In 1935–1936, he found that for certain systems of quaternary quadratic forms, the number of representations of integers by the system satisfied the recursion (1.4.5), in analogy with binary quadratic forms. He published a conjecture on this subject in 1940 [Hec40, Satz 53, p. 100]: that the weighted representation numbers satisfy the Hecke recursion, connecting coefficients to operators on theta series, and further that the columns in a *composition table* always result in linearly independent theta series. He verified the conjecture up to prime level $q < 37$, but was not able to prove this recursion using his methods of complex analysis (see his letter [Bra41, Footnote 1]).

The arithmetic part of these conjectures was investigated by Heinrich Brandt (1886–1954) in the quaternionic context—and so the weave of our narrative is further tightly sewn. Preceding Hecke's work, and inspired by Gauss composition of binary quadratic forms as the product of classes of ideals in a quadratic field, Brandt had earlier considered a generalization to quaternary quadratic forms and the product of classes of ideals in a quaternion algebra [Bra28]: he was only able to define a partially defined product, and so he coined the term *groupoid* for such a structure [Bra40]. He then considered the combinatorial problem of counting the ways of factoring an ideal into prime ideals, according to their classes. In this way, he recorded these counts in a matrix $T(n)$ for each positive integer n, and he proved strikingly (sketched in 1941 [Bra41], dated 1939, and proved completely in 1943 [Bra43]) that the matrices $T(n)$ satisfy Hecke's recursion (1.4.5). To read more on the life and work of Brandt, see Hoehnke–Knus [HK2004]. Today we call the matrices $T(n)$ *Brandt matrices*, and for certain purposes, they are still the most convenient way to get ahold of spaces of modular forms.

Martin Eichler (1912–1992) wrote his thesis [Eic36] under the supervision of Brandt on quaternion orders over the integers, in particular studying the orders that now bear his name. Later he continued the grand synthesis of modular forms, quadratic forms, and quaternion algebras, viewing in generality the orthogonal group of a quadratic form as acting via automorphic transformations [Eic53]. In this vein, he formulated his *basis problem* (arising from the conjecture of Hecke) which sought to understand explicitly the span of quaternionic theta series among classical modular forms, giving a *correspondence* between systems of Hecke eigenvalues appearing in the quaternionic and classical context. He answered the basis problem in affirmative for the case of prime level in 1955 [Eic56a] and then for squarefree level [Eic56b, Eic58, Eic73]. For more on Eichler's basis problem and its history, see Hijikata–Pizer–Shemanske [HPS89a].

Having come to recent history, our account now becomes much more abbreviated: we provide further commentary *in situ* in remarks in the rest of this text, and we conclude with just a few highlights. In the 1950s and 1960s, there was subtantial work done in understanding zeta functions of certain varieties arising from quaternion algebras over totally real number fields. For example, Eichler's correspondence was generalized to totally real fields by Shimizu [Shz65]. Shimura embarked on a deep and systematic study of arithmetic groups obtained from indefinite quaternion algebras over totally real fields, including both the arithmetic Fuchsian groups of Poincaré, Fricke, and Klein, and the generalization of the modular group to totally real fields

studied by Hilbert. In addition to understanding their zeta functions, he also formulated a general theory of complex multiplication in terms of automorphic functions; as a consequence, he found the corresponding arithmetic quotients can be defined as an algebraic variety with equations defined over a number field—and so today we refer to *quaternionic Shimura varieties*. For an overview of Shimura's work, see his lectures at the International Congress of Mathematicians in 1978 [Shi80]. As it turns out, quaternion algebras over number fields also give rise to arithmetic manifolds that are not algebraic varieties, and they are quite important in the areas of spectral theory, low-dimensional geometry, and topology—in particular, in Thurston's geometrization program for hyperbolic 3-manifolds and in classifying knots and links.

Just as the Hecke operators determine the coefficients of classical modular forms and Dirichlet L-series, they may be vastly generalized, replacing modular groups by other algebraic groups, such as the group of units in a central simple algebra or the orthogonal group of a quadratic form. Understanding the theory of automorphic forms in this context is a program that continues today: formulated in the language of automorphic representations, and seen as a nonabelian generalization of class field theory, Langlands initiated this program in a letter to Weil in January 1967. It is indeed fitting that an early success of the *Langlands program* [Gel84, B+2003] would be on the subject of quaternion algebras: a generalization of the Eichler–Shimizu correspondence to encompass arbitrary quaternion algebras over number fields was achieved in foundational work by Jacquet–Langlands [JL70] in 1970. For more on the modern arithmetic history of modular forms, see Edixhoven–Van der Geer–Moonen [EGM2008]; Alsina–Bayer [AB2004, Appendices B–C] also give references for further applications of quaternion algebras in arithmetic geometry (in particular, of Shimura curves).

1.5 Conclusion

We have seen how quaternion algebras have threaded mathematical history through to the present day, weaving together advances in algebra, quadratic forms, number theory, geometry, and modular forms. And although our history ends here, the story does not!

Quaternion algebras continue to arise in unexpected ways. In the arithmetic setting, quaternion orders arise as endomorphism rings of supersingular elliptic curves and have been used in proposed post-quantum cryptosystems and digital signature schemes (see for example the overview by Galbraith–Vercauteren [GV2018]). In the field of quantum computation, Parzanchevski–Sarnak [PS2018] have proposed *Super-Golden-Gates* built from certain special quaternion algebras and their arithmetic groups that would give efficient 1-qubit quantum gates. In coding theory, lattices in quaternion algebras (and more generally central simple algebras over number fields) yield space-time codes that achieve high spectral efficiency on wireless channels with two transmit antennas, currently part of certain IEEE standards [BO2013].

Quaternions have also seen a revival in computer graphics, modeling, and animation [HFK94, Sho85]. Indeed, a rotation in $\mathbb{R}^3$ about an axis through the origin can be represented by a 3×3 orthogonal matrix with determinant 1, conveniently encoded in *Euler angles*. However, the matrix representation is redundant, as there are only three

degrees of freedom in such a rotation. Moreover, to compose two rotations requires the product of the two corresponding matrices, which requires 27 multiplications and 18 additions in $\mathbb{R}$. Quaternions, on the other hand, represent this rotation with a 4-tuple, and multiplication of two quaternions takes only 16 multiplications and 12 additions in $\mathbb{R}$ (if done naively). In computer games, quaternion interpolation provides a way to smoothly interpolate between orientations in space—something crucial for fighting Nazi zombies. Quaternions are also vital for attitude control of aircraft and spacecraft [Hans2006]: they avoid the ambiguity that can arise when two rotation axes align, leading to a potentially disastrous loss of control called *gimbal lock*.

In quantum physics, quaternions yield elegant expressions for Lorentz transformations, the basis of the modern theory of relativity [Gir83]. Some physicists are now hoping to find deeper understanding of these principles of quantum physics in terms of quaternions. And so, although much of Hamilton's quaternionic physics fell out of favor long ago, we have come full circle in our elongated historical arc. The enduring role of quaternion algebras as a catalyst for a vast range of mathematical research promises rewards for many years to come.

Exercises

1. Hamilton sought a multiplication $* : \mathbb{R}^3 \times \mathbb{R}^3 \to \mathbb{R}^3$ that preserves length:

$$\|v\|^2 \cdot \|w\|^2 = \|v * w\|^2$$

for $v, w \in \mathbb{R}^3$. Expanding out in terms of coordinates, such a multiplication would imply that the product of the sum of three squares over $\mathbb{R}$ is again the sum of three squares in $\mathbb{R}$. (Such a law holds for the sum of four squares (1.1.6).) Show that such a formula for three squares is impossible as an identity in the polynomial ring in 6 variables over $\mathbb{Z}$. *[Hint: Find a natural number that is the product of two sums of three squares which is not itself the sum of three squares.]*

2. Hamilton originally sought an associative multiplication law on

$$D := \mathbb{R} + \mathbb{R}i + \mathbb{R}j \simeq \mathbb{R}^3$$

where $i^2 = -1$ and every nonzero element of D has a (two-sided) inverse. Show this cannot happen in two (not really different) ways.

   (a) If $ij = a + bi + cj$ with $a, b, c \in \mathbb{R}$, multiply on the left by i and derive a contradiction.

   (b) Show that D is a (left) $\mathbb{C}$-vector space, so D has even dimension as an $\mathbb{R}$-vector space, a contradiction.

3. Show that there is no way to give $\mathbb{R}^3$ the structure of a ring (with 1) in which multiplication respects scalar multiplication by $\mathbb{R}$, i.e.,

$$x \cdot (cy) = c(x \cdot y) = (cx) \cdot y \quad \text{for all } c \in \mathbb{R} \text{ and } x, y \in \mathbb{R}^3$$

and every nonzero element has a (two-sided) inverse, as follows.

(a) Suppose $B := \mathbb{R}^3$ is equipped with a multiplication law that respects scalar multiplication. Show that left multiplication by $\alpha \in B$ is $\mathbb{R}$-linear and α satisfies the characteristic polynomial of this linear map, a polynomial of degree 3.

(b) Now suppose that every nonzero $\alpha \in B$ has an inverse. By consideration of eigenvalues or the minimal polynomial, derive a contradiction. *[Hint: show that the characteristic polynomial has a real eigenvalue, or that every $\alpha \in B$ satisfies a (minimal) polynomial of degree* 1, *and derive a contradiction from either statement.]*

This chapter is licensed under the terms of the Creative Commons Attribution-NonCommercial 4.0 International License (http://creativecommons.org/licenses/by-nc/4.0/), which permits any noncommercial use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license and indicate if changes were made.

The images or other third party material in this chapter are included in the chapter's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the chapter's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.

Part I
Algebra

Chapter 2

Beginnings

In this chapter, we define quaternion algebras over fields by giving a multiplication table, following Hamilton; we then consider the classical application of understanding rotations in $\mathbb{R}^3$.

2.1 ▹ Conventions

Throughout this text (unless otherwise stated), we let F be a (commutative) field with algebraic closure F^{al}.

When G is a group, and $H \subseteq G$ is a subset, we write $H \leq G$ when H is a subgroup and $H \trianglelefteq G$ when H is a normal subgroup; if G is abelian (written multiplicatively), we write $G^2 := \{g^n : g \in G\} \leq G$ for the subgroup of squares.

We suppose throughout that all rings are associative, not necessarily commutative, with multiplicative identity 1, and that ring homomorphisms preserve 1. In particular, a subring of a ring has the same 1. For a ring A, we write $A^\times$ for the multiplicative group of units of A. An **algebra** over the field F is a ring Bequipped with a homomorphism $F \to B$ such that the image of F lies in the **center** $Z(B)$ of B, defined by

$$Z(B) := \{\alpha \in B : \alpha\beta = \beta\alpha \text{ for all } \beta \in B\}; \tag{2.1.1}$$

if $Z(B) = F$, we say B is **central** (as an F-algebra). We write $\mathrm{M}_n(F)$ for the F-algebra of $n \times n$-matrices with entries in F.

One may profitably think of an F-algebra as being an F-vector space that is also compatibly a ring. If the F-algebra B is not the zero ring, then its structure map $F \to B$ is necessarily injective (since 1 maps to 1) and we identify F with its image; keeping track of the structure map just litters notation. The **dimension** $\dim_F B$ of an F-algebra B is its dimension as an F-vector space.

A **homomorphism** of F-algebras is a ring homomorphism which restricts to the identity on F. An F-algebra homomorphism is necessarily F-linear. An F-algebra homomorphism $B \to B$ is called an **endomorphism**. By convention (and as usual for functions), endomorphisms act on the left. An invertible F-algebra homomorphism $B \xrightarrow{\sim} B'$ is called an **isomorphism**, and an invertible endomorphism is an **automorphism**.

© The Author(s) 2021

J. Voight, *Quaternion Algebras*, Graduate Texts in Mathematics 288,
https://doi.org/10.1007/978-3-030-56694-4_2

The set of automorphisms of B forms a group, which we write as $\operatorname{Aut}(B)$—these maps are necessarily F-linear, but we do not include this in the notation. reserve the notation $\operatorname{End}_F(V)$ for the ring of F-linear endomorphisms of the F-vector space V, and $\operatorname{Aut}_F(V)$ for the group of F-linear automorphisms of V; in particular, $\operatorname{End}_F(B) \sim \mathrm{M}_n(F)$ if $n = \dim_F B$.

Remark 2.1.2. Throughout, whenever we define a homomorphism of objects, we adopt the (categorical) convention extending this to the terms *endomorphism* (homomorphism with equal domain and codomain), *isomorphism* (invertible homomorphism), and *automorphism* (invertible endomorphism).

A **division ring** (also called a **skew field**) is a ring D in which every nonzero element has a (two-sided) inverse, i.e., $D \smallsetminus \{0\}$ is a group under multiplication. A **division algebra** is an algebra that is a division ring.

2.2 ▹ Quaternion algebras

In this section, we define quaternion algebras in a direct way, via generators and relations. Throughout the rest of this chapter, suppose that $\operatorname{char} F \neq 2$; the case $\operatorname{char} F = 2$ is treated in Chapter 6.

Definition 2.2.1. An algebra B over F is a **quaternion algebra** if there exist $i, j \in B$ such that $1, i, j, ij$ is an F-basis for B and

$$i^2 = a, \quad j^2 = b, \quad \text{and} \quad ji = -ij \tag{2.2.2}$$

for some $a, b \in F^\times$.

The entire multiplication table for a quaternion algebra is determined by the multiplication rules (2.2.2), linearity, and associativity: for example,

$$(ij)^2 = (ij)(ij) = i(ji)j = i(-ij)j = -(i^2)(j^2) = -ab$$

and $j(ij) = (-ij)j = -bi$. Conversely, given $a, b \in F^\times$, one can write down the unique possible associative multiplication table on the basis $1, i, j, k$ compatible with (2.2.2), and then verify independently that it is associative (Exercise 2.1). Accordingly, for $a, b \in F^\times$, we define $\left(\dfrac{a, b}{F}\right)$ to be the quaternion algebra over F with F-basis $1, i, j, ij$ subject to the multiplication (2.2.2); we will also write $(a, b \mid F)$ when convenient for formatting. By definition, we have $\dim_F(a, b \mid F) = 4$.

The map which interchanges i and j gives an isomorphism $\left(\dfrac{a, b}{F}\right) \simeq \left(\dfrac{b, a}{F}\right)$, so Definition 2.2.1 is symmetric in a, b. The elements a, b are far from unique in determining the isomorphism class of a quaternion algebra: see Exercise 2.4.

If $K \supseteq F$ is a field extension of F, then there is a canonical isomorphism

$$\left(\frac{a, b}{F}\right) \otimes_F K \simeq \left(\frac{a, b}{K}\right)$$

extending scalars (same basis, but now spanning a K-vector space), so Definition 2.2.1 behaves well with respect to inclusion of fields.

Example 2.2.3. The $\mathbb{R}$-algebra $\mathbb{H} := \left(\dfrac{-1,-1}{\mathbb{R}}\right)$ is the ring of quaternions over the real numbers, discovered by Hamilton; we call $\mathbb{H}$ the ring of **(real) Hamiltonians** (also known as **Hamilton's quaternions**).

Example 2.2.4. The ring $\mathrm{M}_2(F)$ of 2×2-matrices with coefficients in F is a quaternion algebra over F: there is an isomorphism $\left(\dfrac{1,1}{F}\right) \xrightarrow{\sim} \mathrm{M}_2(F)$ of F-algebras induced by

$$i \mapsto \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}, \; j \mapsto \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}.$$

If $F = F^{\mathrm{al}}$ is algebraically closed and B is a quaternion algebra over F, then necessarily $B \simeq \mathrm{M}_2(F)$ (Exercise 2.4). Consequently, every quaternion algebra B over F has $B \otimes_F F^{\mathrm{al}} \simeq \mathrm{M}_2(F^{\mathrm{al}})$.

A quaternion algebra B is generated by the elements i, j by definition (2.2.2). However, exhibiting an algebra by generators and relations (instead of by a multiplication table) can be a bit subtle, as the dimension of such an algebra is not a priori clear. But working with presentations is quite useful; and at least for quaternion algebras, we can think in these terms as follows.

Lemma 2.2.5. *An F-algebra B is a quaternion algebra if and only if there exist nonzero $i, j \in B$ that generate B as an F-algebra and satisfy*

$$i^2 = a, \; j^2 = b, \; \text{ and } \; ij = -ji \tag{2.2.6}$$

with $a, b \in F^\times$.

In other words, once the relations (2.2.6) are satisfied for generators i, j, then automatically B has dimension 4 as an F-vector space, with F-basis $1, i, j, ij$.

Proof. It is necessary and sufficient to prove that the elements $1, i, j, ij$ are linearly independent. Suppose that $\alpha = t + xi + yj + zij = 0$ with $t, x, y, z \in F$. Using the relations given, we compute that

$$0 = i(\alpha i + i\alpha) = 2a(t + xi).$$

Since $\operatorname{char} F \neq 2$ and $a \neq 0$, we conclude that $t + xi = 0$. Repeating with j and ij, we similarly find that $t + yj = t + zij = 0$. Thus

$$\alpha - (t + xi) - (t + yj) - (t + zij) = -2t = 0.$$

Since i, j are nonzero, B is not the zero ring, so $1 \neq 0$; thus $t = 0$ and so $xi = yj = zij = 0$. Finally, if $x \neq 0$, then $i = 0$ so $i^2 = 0 = a$, impossible; hence $x = 0$. Similarly, $y = z = 0$. □

Accordingly, we will call elements $i, j \in B$ satisfying (2.2.6) **standard generators** for a quaternion algebra B.

Remark 2.2.7. Invertibility of both a and b in F is needed for Lemma 2.2.5: the commutative algebra $B = F[i,j]/(i,j)^2$ is generated by the elements i,j satisfying $i^2 = j^2 = ij = -ji = 0$ but B is *not* a quaternion algebra.

Remark 2.2.8. In light of Lemma 2.2.5, we will often drop the symbol $k = ij$ and reserve it for other use. (In particular, in later sections we will want k to represent other quaternion elements.) If we wish to use this abbreviation, we will assign $k := ij$.

2.3 ▹ Matrix representations

Every quaternion algebra can be viewed as a subalgebra of 2×2-matrices over an at most quadratic extension; this is sometimes taken to be the definition!

Proposition 2.3.1. *Let* $B := \left(\dfrac{a,b}{F}\right)$ *be a quaternion algebra over* F *and let* $F(\sqrt{a})$ *be a splitting field over* F *for the polynomial* $x^2 - a$, *with root* $\sqrt{a} \in F(\sqrt{a})$. *Then the map*

$$\begin{aligned} \lambda\colon B &\to \mathrm{M}_2(F(\sqrt{a})) \\ i, j &\mapsto \begin{pmatrix} \sqrt{a} & 0 \\ 0 & -\sqrt{a} \end{pmatrix}, \begin{pmatrix} 0 & b \\ 1 & 0 \end{pmatrix} \\ t + xi + yj + zij &\mapsto \begin{pmatrix} t + x\sqrt{a} & b(y + z\sqrt{a}) \\ y - z\sqrt{a} & t - x\sqrt{a} \end{pmatrix} \end{aligned} \tag{2.3.2}$$

is an injective F*-algebra homomorphism and an isomorphism onto its image.*

Proof. Injectivity follows by checking $\ker \lambda = \{0\}$ on matrix entries, and the homomorphism property can be verified directly, checking the multiplication table (Exercise 2.10). □

Remark 2.3.3. Proposition 2.3.1 can be turned around to assert the existence of quaternion algebras: one can check that the set

$$\left\{ \begin{pmatrix} t + x\sqrt{a} & b(y + z\sqrt{a}) \\ y - z\sqrt{a} & t - x\sqrt{a} \end{pmatrix} : t, x, y, z \in F \right\} \subseteq \mathrm{M}_2(F(\sqrt{a}))$$

is an F-vector subspace of dimension 4, closed under multiplication, with the matrices $\lambda(i), \lambda(j)$ satisfying the defining relations (2.2.2).

2.3.4. If $a \notin F^{\times 2}$, then $K = F(\sqrt{a}) \supseteq F$ is a quadratic extension of F. Let $\mathrm{Gal}(K\,|\,F) = \mathrm{Aut}_F(K) \simeq \mathbb{Z}/2\mathbb{Z}$ be the Galois group of K over F and let $\sigma \in \mathrm{Gal}(K\,|\,F)$ be the nontrivial element. Then we can rewrite the image $\lambda(B)$ in (2.3.2) as

$$\lambda(B) = \left\{ \begin{pmatrix} u & bv \\ \sigma(v) & \sigma(u) \end{pmatrix} : u, v \in K \right\} \subset \mathrm{M}_2(K). \tag{2.3.5}$$

Corollary 2.3.6. *We have an isomorphism*

$$\begin{aligned} \left(\frac{1,b}{F}\right) &\xrightarrow{\sim} \mathrm{M}_2(F) \\ i,j &\mapsto \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}, \begin{pmatrix} 0 & b \\ 1 & 0 \end{pmatrix} \end{aligned} \tag{2.3.7}$$

Proof. Specializing Proposition 2.3.1, we see the map is an injective F-algebra homomorphism, so since $\dim_F B = \dim_F M_2(F) = 4$, the map is also surjective. □

The provenance of the map (2.3.2) is itself important, so we now pursue another (more natural) proof of Proposition 2.3.1.

2.3.8. Let

$$K := F[i] = F \oplus Fi \simeq F[x]/(x^2 - a)$$

be the (commutative) F-algebra generated by i. Suppose first that K is a field (so $a \notin F^{\times 2}$): then $K \simeq F(\sqrt{a})$ is a quadratic field extension of F. The algebra B has the structure of a right K-vector space of dimension 2, with basis $1, j$: explicitly,

$$\alpha = t + xi + yj + zij = (t + xi) + j(y - zi) \in K \oplus jK$$

for all $\alpha \in B$, so $B = K \oplus jK$. We then define the **left regular representation** of B over K by

$$\begin{aligned} \lambda \colon B &\to \mathrm{End}_K(B) \\ \alpha &\mapsto (\lambda_\alpha \colon \beta \mapsto \alpha\beta). \end{aligned} \tag{2.3.9}$$

Each map λ_α is indeed a K-linear endomorphism in B (considered as a right K-vector space) by associativity in B: for all $\alpha, \beta \in B$ and $w \in K$,

$$\lambda_\alpha(\beta w) = \alpha(\beta w) = (\alpha\beta)w = \lambda_\alpha(\beta)w.$$

Similarly, λ is an F-algebra homomorphism: for all $\alpha, \beta, \nu \in B$

$$\lambda_{\alpha\beta}(\nu) = (\alpha\beta)\nu = \alpha(\beta(\nu)) = (\lambda_\alpha\lambda_\beta)(\nu)$$

reading functions from right to left as usual. The map λ is injective (λ is a **faithful** representation) since $\lambda_\alpha = 0$ implies $\lambda_\alpha(1) = \alpha = 0$.

In the basis $1, j$ we have $\mathrm{End}_K(B) \simeq \mathrm{M}_2(K)$, and λ is given by

$$i \mapsto \lambda_i = \begin{pmatrix} i & 0 \\ 0 & -i \end{pmatrix}, \quad j \mapsto \lambda_j = \begin{pmatrix} 0 & b \\ 1 & 0 \end{pmatrix}; \tag{2.3.10}$$

these matrices act on column vectors on the left. We then recognize the map λ given in (2.3.2).

If K is not a field, then $K \simeq F \times F$, and we repeat the above argument but with B a free module of rank 2 over K; then projecting onto one of the factors (choosing $\sqrt{a} \in F$) gives the map λ, which is still injective and therefore induces an F-algebra isomorphism $B \simeq \mathrm{M}_2(F)$.

Remark 2.3.11. In Proposition 2.3.1, B acts on columns on the left; if instead, one wishes to have B act on the right on rows, give B the structure of a left K-vector space and define accordingly the right regular representation instead (taking care about the order of multiplication).

2.3.12. In some circumstances, it can be notationally convenient to consider variants of the injection (2.3.2): for example

$$\begin{aligned} B &\to \mathrm{M}_2(F(\sqrt{a})) \\ t + xi + yj + zij &\mapsto \begin{pmatrix} t + x\sqrt{a} & y + z\sqrt{a} \\ b(y - z\sqrt{a}) & t - x\sqrt{a} \end{pmatrix} \end{aligned} \tag{2.3.13}$$

is obtained by taking the basis $1, b^{-1}j$, equivalently postcomposing by $\begin{pmatrix} 1 & 0 \\ 0 & b \end{pmatrix}$. See also Exercise 2.12.

Remark 2.3.14. The left regular representation 2.3.8 is not the only way to embed B as a subalgebra of 2×2-matrices. Indeed, the "splitting" of quaternion algebras in this way, in particular the question of whether or not $B \simeq \mathrm{M}_2(F)$, is a theme that will reappear throughout this text. For a preview, see Main Theorem 5.4.4.

2.3.15. Thinking of a quaternion algebra as in 2.3.8 as a right K-vector space suggests notation for quaternion algebras that is also useful: for a peek, see 6.1.5.

2.4 ▹ Rotations

To conclude this chapter, we return to Hamilton's original design: quaternions model rotations in 3-dimensional space. This development is not only historically important but it also previews many aspects of the general theory of quaternion algebras over fields. In this section, we follow Hamilton and take $k := ij$.

Proposition 2.3.1 provides an $\mathbb{R}$-algebra embedding

$$\begin{aligned} \lambda \colon \mathbb{H} &\hookrightarrow \mathrm{End}_{\mathbb{C}}(\mathbb{H}) \simeq \mathrm{M}_2(\mathbb{C}) \\ t + xi + yj + zk = u + j\overline{v} &\mapsto \begin{pmatrix} t + xi & -y - zi \\ y - zi & t - xi \end{pmatrix} = \begin{pmatrix} u & -v \\ \overline{v} & \overline{u} \end{pmatrix} \end{aligned} \tag{2.4.1}$$

where $u := t + xi$ and $v := y + zi$ and $\overline{}$ denotes complex conjugation. (The abuse of notation, taking $i \in \mathbb{H}$ as well as $i \in \mathbb{C}$ is harmless: we may think of $\mathbb{C} \subset \mathbb{H}$.) We have

$$\det \begin{pmatrix} u & -v \\ \overline{v} & \overline{u} \end{pmatrix} = |u|^2 + |v|^2 = t^2 + x^2 + y^2 + z^2,$$

thus $\mathbb{H}^\times = \mathbb{H} \smallsetminus \{0\}$. If preferred, see (2.3.13) to obtain matrices of the form $\begin{pmatrix} u & v \\ -\overline{v} & \overline{u} \end{pmatrix}$ instead.

2.4.2. We define the subgroup of **unit Hamiltonians** as

$$\mathbb{H}^1 := \{t + xi + yj + zk \in \mathbb{H} : t^2 + x^2 + y^2 + z^2 = 1\}.$$

(In some contexts, one also writes $\mathrm{GL}_1(\mathbb{H}) = \mathbb{H}^\times$ and $\mathrm{SL}_1(\mathbb{H}) = \mathbb{H}^1$.)

As a set, the unit Hamiltonians are naturally identified with the 3-sphere in $\mathbb{R}^4$. As groups, we have an isomorphism $\mathbb{H}^1 \simeq \mathrm{SU}(2)$ with the special unitary group of rank 2, defined by

$$\mathrm{SU}(n) := \{A \in \mathrm{SL}_n(\mathbb{C}) : A^* = A^{-1}\} = \{A \in \mathrm{SL}_n(\mathbb{C}) : JA = \overline{A}J\} \tag{2.4.3}$$

where $A^* = \overline{A}^{\mathsf{t}}$ is the (complex) conjugate transpose of A and $J := \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}$ is the image of $j \in \mathbb{H}^1$.

Definition 2.4.4. Let $\alpha \in \mathbb{H}$. We say α is **real** if $\alpha \in \mathbb{R}$, and we say α is **pure** (or **imaginary**) if $\alpha \in \mathbb{R}i + \mathbb{R}j + \mathbb{R}k$.

2.4.5. Just as for the complex numbers, every element of $\mathbb{H}$ is the sum of its real part and its pure (imaginary) part. And just like complex conjugation, we define a **(quaternion) conjugation** map

$$\begin{aligned} \overline{} : \mathbb{H} &\to \mathbb{H} \\ \alpha = t + (xi + yj + zk) &\mapsto \overline{\alpha} = t - (xi + yj + zk) \end{aligned} \tag{2.4.6}$$

by negating the imaginary part. We compute that

$$\begin{aligned} \alpha + \overline{\alpha} &= \mathrm{tr}(\lambda(\alpha)) = 2t \\ \|\alpha\|^2 &:= \det(\lambda(\alpha)) = \alpha\overline{\alpha} = \overline{\alpha}\alpha = t^2 + x^2 + y^2 + z^2. \end{aligned} \tag{2.4.7}$$

The notation $\|\,\|^2$ is used to indicate that it agrees the usual square norm on $\mathbb{H} \simeq \mathbb{R}^4$.

The conjugate transpose map on $\mathrm{M}_2(\mathbb{C})$ restricts to quaternion conjugation on the image of $\mathbb{H}$ in (2.4.1), also known as **adjugation**

$$\alpha = \begin{pmatrix} u & -v \\ \overline{v} & \overline{u} \end{pmatrix} \mapsto \lambda(\overline{\alpha}) = \begin{pmatrix} \overline{u} & v \\ -\overline{v} & u \end{pmatrix}.$$

Thus the elements $\alpha \in \mathbb{H}$ such that $\lambda(\overline{\alpha}) = \lambda(\alpha)$ (i.e., $A^* = A$, and we say A is **Hermitian**), are exactly the scalar (real) matrices; and those that are **skew-Hermitian**, i.e., $A^* = -A$, are exactly the pure quaternions. The conjugation map plays a crucial role for quaternion algebras and is the subject of the next chapter (Chapter 3), where to avoid confusion with other notions of conjugation we refer to it as the *standard involution*.

2.4.8. Let

$$\mathbb{H}^0 := \{v = xi + yj + zk \in \mathbb{H} : x, y, z \in \mathbb{R}\} \simeq \mathbb{R}^3$$

be the set of pure Hamiltonians, the three-dimensional real space on which we will soon see that the (unit) Hamiltonians act by rotations. (The reader should not confuse $v \in \mathbb{H}^0$ with v the entry of a 2×2-matrix in a local instantiation above.) For $v \in \mathbb{H}^0 \simeq \mathbb{R}^3$,

$$\|v\|^2 = x^2 + y^2 + z^2 = \det(\lambda(v)), \tag{2.4.9}$$

and from (2.4.1),

$$\mathbb{H}^0 = \{v \in \mathbb{H} : \operatorname{tr}(\lambda(v)) = v + \overline{v} = 0\}.$$

We again see that $\overline{v} = -v$ for $v \in \mathbb{H}^0$.

The set $\mathbb{H}^0$ is not closed under multiplication: if $v, w \in \mathbb{H}^0$, then

$$vw = -v \cdot w + v \times w \tag{2.4.10}$$

where $v \cdot w$ is the dot product on $\mathbb{R}^3$ and $v \times w \in \mathbb{H}^0$ is the **cross product**, defined as the determinant

$$v \times w = \det \begin{pmatrix} i & j & k \\ v_1 & v_2 & v_3 \\ w_1 & w_2 & w_3 \end{pmatrix} \tag{2.4.11}$$

where $v = v_1 i + v_2 j + v_3 k$ and $w = w_1 i + w_2 j + w_3 k$, so

$$v \cdot w = v_1 w_1 + v_2 w_2 + v_3 w_3$$

and

$$v \times w = (v_2 w_3 - v_3 w_2) i + (v_3 w_1 - v_1 w_3) j + (v_1 w_2 - v_2 w_1) k.$$

The formula (2.4.10) is striking: it contains three different kinds of ‘multiplications’!

Lemma 2.4.12. *For all $v, w \in \mathbb{H}^0$, the following statements hold.*

(a) *$vw \in \mathbb{H}^0$ if and only if v, w are orthogonal.*

(b) $v^2 = -\|v\|^2 \in \mathbb{R}$.

(c) *$wv = -vw$ if and only if v, w are orthogonal.*

Proof. Apply (2.4.10). □

2.4.13. The group $\mathbb{H}^1$ acts on our three-dimensional space $\mathbb{H}^0$ (on the left) by conjugation:

$$\begin{aligned} \mathbb{H}^1 \circlearrowright \mathbb{H}^0 &\to \mathbb{H}^0 \\ v &\mapsto \alpha v \alpha^{-1}; \end{aligned} \tag{2.4.14}$$

indeed, $\operatorname{tr}(\lambda(\alpha v \alpha^{-1})) = \operatorname{tr}(\lambda(v)) = 0$ by properties of the trace, so $\alpha v \alpha^{-1} \in \mathbb{H}^0$. Or

$$\mathbb{H}^0 = \{v \in \mathbb{H} : v^2 \in \mathbb{R}_{\leq 0}\}$$

and this latter set is visibly stable under conjugation. The representation (2.4.14) is called the **adjoint representation**.

2.4.15. Let $\alpha \in \mathbb{H}^1 \smallsetminus \{\pm 1\}$. Then there exists a unique $\theta \in (0, \pi)$ such that

$$\alpha = t + xi + yj + zk = \cos\theta + (\sin\theta) I(\alpha) \tag{2.4.16}$$

where $I(\alpha)$ is pure and $\|I(\alpha)\| = 1$: to be precise, we take θ such that $\cos\theta = t$ and

$$I(\alpha) := \frac{xi + yj + zk}{|\sin\theta|}.$$

We call $I(\alpha)$ the **axis** of α, and observe that $I(\alpha)^2 = -1$.

Remark 2.4.17. In analogy with Euler's formula, we can write (2.4.16) as

$$\alpha = \exp(I(\alpha)\theta).$$

We are now prepared to identify this action by quaternions with rotations. As usual, let

$$\mathrm{O}(n) := \{A \in \mathrm{M}_n(\mathbb{R}) : A^{\mathsf{t}} = A^{-1}\}$$

be the orthogonal group of $\mathbb{R}^n$ (preserving the standard inner product), and let

$$\mathrm{SO}(n) := \{A \in \mathrm{O}(n) : \det(A) = 1\} \trianglelefteq \mathrm{O}(n)$$

to be the special orthogonal group of rotations of $\mathbb{R}^n$, a normal subgroup of index 2 fitting into the exact sequence

$$1 \to \mathrm{SO}(n) \to \mathrm{O}(n) \xrightarrow{\det} \{\pm 1\} \to 1.$$

Proposition 2.4.18. *$\mathbb{H}^1$ acts by rotation on $\mathbb{H}^0 \simeq \mathbb{R}^3$ via conjugation (2.4.14): specifically, α acts by rotation through the angle 2θ about the axis $I(\alpha)$.*

Proof. Let $\alpha \in \mathbb{H}^1 \smallsetminus \{\pm 1\}$. Then for all $v \in \mathbb{H}^0$,

$$\|\alpha v \alpha^{-1}\|^2 = \|v\|^2$$

by (2.4.9), so α acts by a matrix belonging to O(3).

But we can be more precise. Let $j' \in \mathbb{H}^0$ be a unit vector orthogonal to $i' = I(\alpha)$. Then $(i')^2 = (j')^2 = -1$ by Lemma 2.4.12(b) and $j'i' = -i'j'$ by Lemma 2.4.12(c), so without loss of generality we may suppose that $I(\alpha) = i$ and $j' = j$. Thus $\alpha = t + xi = \cos\theta + (\sin\theta)i$ with $t^2 + x^2 = \cos^2\theta + \sin^2\theta = 1$, and $\alpha^{-1} = t - xi$.

We have $\alpha i \alpha^{-1} = i$, and

$$\begin{aligned} \alpha j \alpha^{-1} &= (t + xi)j(t - xi) = (t + xi)(t + xi)j \\ &= ((t^2 - x^2) + 2txi)j = (\cos 2\theta)j + (\sin 2\theta)k \end{aligned} \tag{2.4.19}$$

by the double angle formula. Consequently,

$$\alpha k \alpha^{-1} = i(\alpha j \alpha^{-1}) = (-\sin 2\theta)j + (\cos 2\theta)k$$

so the matrix of α in the basis i, j, k is

$$A = \begin{pmatrix} 1 & 0 & 0 \\ 0 & \cos 2\theta & -\sin 2\theta \\ 0 & \sin 2\theta & \cos 2\theta \end{pmatrix}, \tag{2.4.20}$$

a (counterclockwise) rotation (determinant 1) through the angle 2θ about i. ∎

Corollary 2.4.21. *The action* (2.4.13) *defines a group homomorphism* $\mathbb{H}^1 \to \mathrm{SO}(3)$*, fitting into an exact sequence*

$$1 \to \{\pm 1\} \to \mathbb{H}^1 \to \mathrm{SO}(3) \to 1.$$

Proof. The map $\mathbb{H}^1 \to \mathrm{SO}(3)$ is surjective, since every element of $\mathrm{SO}(3)$ is rotation about some axis (Exercise 2.15). If α belongs to the kernel, then $\alpha = \cos\theta + (\sin\theta)I(\alpha)$ must have $\sin\theta = 0$ so $\alpha = \pm 1$. □

2.4.22. The matrix representation of $\mathbb{H}$ in section 2.4 extends to a matrix representation of $\mathbb{H} \otimes_{\mathbb{R}} \mathbb{C}$, and this representation and its connection to unitary matrices is still used widely in quantum mechanics. In the embedding with

$$i \mapsto \begin{pmatrix} i & 0 \\ 0 & -i \end{pmatrix}, \quad -j \mapsto \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}, \quad -k \mapsto \begin{pmatrix} 0 & i \\ i & 0 \end{pmatrix}$$

whose images are unitary matrices, we multiply by $-i$ to obtain Hermitian matrices

$$\sigma_z := \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}, \quad \sigma_y := \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix}, \quad \sigma_x := \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$$

where $\sigma_x, \sigma_y, \sigma_z$ are the famous **Pauli spin matrices**. Because of this application to the spin (a kind of angular momentum) of an electron in particle physics, the group $\mathbb{H}^1$ also goes by the name $\mathbb{H}^1 \simeq \mathrm{Spin}(3)$.

The extra bit of information conveyed by spin can also be seen by the "belt trick" [Hans2006, Chapter 2].

2.4.23. We conclude with one final observation, returning to the formula (2.4.10). There is another way to mix the dot product and cross product (2.4.11) in $\mathbb{H}$: we define the **scalar triple product**

$$\begin{aligned} \mathbb{H} \times \mathbb{H} \times \mathbb{H} &\to \mathbb{R} \\ (u, v, w) &\mapsto u \cdot (v \times w). \end{aligned} \tag{2.4.24}$$

Amusingly, this gives a way to "multiply" *triples* of triples! The map (2.4.24) defines an alternating, trilinear form (Exercise 2.19). If $u, v, w \in \mathbb{H}^0$, then the scalar triple product is a determinant

$$u \cdot (v \times w) = \det \begin{pmatrix} u_1 & u_2 & u_3 \\ v_1 & v_2 & v_3 \\ w_1 & w_2 & w_3 \end{pmatrix}$$

and $|u \cdot (v \times w)|$ is the volume of a parallelepiped in $\mathbb{R}^3$ whose sides are given by u, v, w.

Exercises

Let F be a field with $\operatorname{char} F \neq 2$.

▹ 1. Show that a (not necessarily associative) F-algebra is associative if and only if the associative law holds on a basis, and then check that the multiplication table implied by (2.2.2) is associative.

2. Show that if B is an F-algebra generated by $i, j \in B$ and $1, i, j$ are linearly dependent, then B is commutative.

3. Verify directly that the map $\left(\dfrac{1,1}{F}\right) \xrightarrow{\sim} \mathrm{M}_2(F)$ in Example 2.2.4 is an isomorphism of F-algebras.

▹ 4. Let $a, b \in F^\times$.

   (a) Show that $\left(\dfrac{a,b}{F}\right) \simeq \left(\dfrac{a,-ab}{F}\right) \simeq \left(\dfrac{b,-ab}{F}\right)$.

   (b) Show that if $c, d \in F^\times$ then $\left(\dfrac{a,b}{F}\right) \simeq \left(\dfrac{ac^2,bd^2}{F}\right)$. Conclude that if $F^\times/F^{\times 2}$ is finite, then there are only finitely many isomorphism classes of quaternion algebras over F, and in particular that if $F^{\times 2} = F^\times$ then there is only one isomorphism class $\left(\dfrac{1,1}{F}\right) \simeq \mathrm{M}_2(F)$. [The converse is not true, see Exercise 3.16.]

   (c) Show that if $B = \left(\dfrac{a,b}{\mathbb{R}}\right)$ is a quaternion algebra over $\mathbb{R}$, then $B \simeq \mathrm{M}_2(\mathbb{R})$ or $B \simeq \mathbb{H}$, the latter occurring if and only if $a < 0$ and $b < 0$. Conclude that if B is a division quaternion algebra over $\mathbb{R}$, then $B \simeq \mathbb{H}$.

   (d) Let B be a quaternion algebra over F. Show that $B \otimes_F F^{\mathrm{al}} \simeq \mathrm{M}_2(F^{\mathrm{al}})$, where F^{al} is an algebraic closure of F.

   (e) Refine part (d) as follows. A field $K \supseteq F$ is a **splitting field** for B if $B \otimes_F K \simeq \mathrm{M}_2(K)$. Show that B has a splitting field K with $[K : F] \leq 2$.

5. Let $B = \left(\dfrac{a,b}{F}\right)$ be a quaternion algebra over F. Let $i' \in B \smallsetminus F$ satisfy $(i')^2 = a' \in F^\times$. Show that there exists $b' \in F^\times$ and an isomorphism $B \simeq \left(\dfrac{a',b'}{F}\right)$ (under which i' maps to the first standard generator).

6. Use the quaternion algebra $B = \left(\dfrac{-1,-1}{F}\right)$, multiplicativity of the determinant, and the left regular representation (2.3.2) to show that if two elements of F can be written as the sum of four squares, then so too can their product (a discovery of Euler in 1748). [In Chapter 3, this statement will follow immediately from the multiplicativity of the *reduced norm* on B; here, the formula is derived easily from multiplicativity of the determinant.]

▹ 7. Let B be an F-algebra. Show that if B is a quaternion algebra over F, then B is central.

▹ 8. Let A, B be F-algebras, and let $\phi\colon A \to B$ be a surjective F-algebra homomorphism. Show that ϕ restricts to an F-algebra homomorphism $Z(A) \to Z(B)$.

▹ 9. Prove the following partial generalization of Exercise 2.4(b). Let B be a finite-dimensional algebra over F.

(a) Show that every element $\alpha \in B$ satisfies a unique monic polynomial of smallest degree with coefficients in F.
(b) Suppose that $B = D$ is a division algebra. Show that the minimal polynomial of $\alpha \in D$ is irreducible over F. Conclude that if $F = F^{\mathrm{al}}$ is algebraically closed, then $D = F$.

▹ 10. Prove Proposition 2.3.1: show directly that the map

$$\begin{aligned} \lambda\colon B &\to \mathrm{M}_2(F(\sqrt{a})) \\ i, j &\mapsto \begin{pmatrix} \sqrt{a} & 0 \\ 0 & -\sqrt{a} \end{pmatrix}, \begin{pmatrix} 0 & b \\ 1 & 0 \end{pmatrix} \end{aligned}$$

extends uniquely to an injective F-algebra homomorphism. *[Hint: check that the relations are satisfied.]*

11. a) Show explicitly that every quaternion algebra $B = (a, b \mid F)$ is isomorphic to an F-subalgebra of $\mathrm{M}_4(F)$ via the left (or right) regular representation over F: write down 4×4-matrices representing i and j and verify that the relations $i^2 = a$, $j^2 = b$, $ji = -ij$ hold for these matrices. Note the 2×2-block structure of these matrices.

b) With respect to a suitable such embedding in (a) for $B = \mathbb{H}$, verify that the quaternionic conjugation map $\alpha \mapsto \overline{\alpha}$ is the matrix transpose, and the matrix determinant is the *square* of the norm $\|\alpha\|^2 = \alpha\overline{\alpha}$.

12. In certain circumstances, one may not want to "play favorites" in the left regular representation (Proposition 2.3.1) and so involve i and j on more equal footing. To this end, show that the map

$$\begin{aligned} B = \left(\frac{a, b}{F}\right) &\to \mathrm{M}_2(F(\sqrt{a}, \sqrt{b})) \\ i, j &\mapsto \begin{pmatrix} \sqrt{a} & 0 \\ 0 & -\sqrt{a} \end{pmatrix}, \begin{pmatrix} 0 & \sqrt{b} \\ \sqrt{b} & 0 \end{pmatrix} \end{aligned} \tag{2.4.25}$$

is an injective F-algebra isomorphism. How is it related to the left regular representation?

13. Let $B = (a, b \mid F)$ be a quaternion algebra over F. For a nonzero element $\alpha = t + xi + yj + zk \in B$, show that the following are equivalent:

(i) $t = 0$; and
(ii) $\alpha \notin F$ and $\alpha^2 \in F$.

[So the notion of "pure quaternion" is not tethered to a particular basis.]

14. Verify that (2.3.7) is an isomorphism of F-algebras, and interpret this map as arising from the left regular representation via a map $B \hookrightarrow \mathrm{M}_2(F \times F) \to \mathrm{M}_2(F)$.

15. Show that every rotation $A \in \mathrm{SO}(3)$ fixes an axis. *[Hint: Consider the eigenvalues of A.]*

16. For $v \in \mathbb{H}^0$ and $\beta \in \mathbb{H}^0 \smallsetminus \{0\}$, consider the map $v \mapsto \beta^{-1}\overline{v}\beta = -\beta^{-1}v\beta \in \mathbb{H}^0$. Show that this map is the reflection across the plane $\{w \in \mathbb{H}^0 : \beta w = 0\}$. (For example, taking $\beta = i$, the map is $xi + yj + zk \mapsto -xi + yj + zk$.)

▹ 17. In Corollary 2.4.21, we showed that $\mathrm{SU}(2) \simeq \mathbb{H}^1$ has a 2-to-1 map to $\mathrm{SO}(3)$, where $\mathbb{H}^1$ acts on $\mathbb{H}^0 \simeq \mathbb{R}^3$ by conjugation: quaternions model rotations in three-dimensional space, with spin. Quaternions also model rotations in four-dimensional space, as follows.

 (a) Show that the map

$$\begin{aligned} (\mathbb{H}^1 \times \mathbb{H}^1) \circlearrowright \mathbb{H} &\to \mathbb{H} \\ x &\mapsto \alpha x \beta^{-1} \end{aligned} \tag{2.4.26}$$

 defines a (left) action of $\mathbb{H}^1 \times \mathbb{H}^1$ on $\mathbb{H} \simeq \mathbb{R}^4$, giving a group homomorphism

$$\phi \colon \mathbb{H}^1 \times \mathbb{H}^1 \to \mathrm{O}(4).$$

 (b) Show that ϕ surjects onto $\mathrm{SO}(4) < \mathrm{O}(4)$. *[Hint: If $A \in \mathrm{SO}(4)$ fixes $1 \in \mathbb{H}$, then A restricted to $\mathbb{H}^0$ is a rotation and so is given by conjugation. More generally, if $A1 = \alpha$, consider $x \mapsto \alpha^{-1}Ax$.]*

 (c) Show that the kernel of ϕ is $\{\pm 1\}$ embedded diagonally, so there is an exact sequence

$$1 \to \{\pm 1\} \to \mathrm{SU}(2) \times \mathrm{SU}(2) \to \mathrm{SO}(4) \to 1.$$

 [More generally, the universal cover of $\mathrm{SO}(n)$ for $n \geq 3$ is a double cover called the **spin group** $\mathrm{Spin}(n)$, and so Corollary 2.4.21 shows that $\mathrm{Spin}(3) \simeq \mathrm{SU}(2)$ and this exercise shows that $\mathrm{Spin}(4) \simeq \mathrm{SU}(2) \times \mathrm{SU}(2)$. For further reading, see e.g. Fulton–Harris [FH91, Lecture 20].]

18. Let $\rho_{u,\theta} \colon \mathbb{R}^3 \to \mathbb{R}^3$ be the counterclockwise rotation by the angle θ about the axis $u \in \mathbb{R}^3 \simeq \mathbb{H}^0$, with $\|u\| = 1$. Prove **Rodrigues's rotation formula**: for all $v \in \mathbb{R}^3$,

$$\rho_{u,\theta}(v) = (\cos\theta)v + (\sin\theta)(u \times v) + (1 - \cos\theta)(u \cdot v)u$$

 where $u \times v$ and $u \cdot v$ are the cross and dot product, respectively.

19. Verify that the map (2.4.24) is a trilinear alternating form on $\mathbb{H}$, i.e., show the form is linear when any two of the three arguments are fixed and zero when two argument are equal.

20. Let B be a quaternion algebra over F and let $\mathrm{M}_2(B)$ be the ring of 2×2-matrices over B. (Be careful in the definition of matrix multiplication: B is

noncommutative!) Consider the Cayley determinant:

$$\begin{aligned} \mathrm{Cdet}\colon\ \mathrm{M}_2(B) &\to B \\ \mathrm{Cdet}\begin{pmatrix} \alpha & \beta \\ \gamma & \delta \end{pmatrix} &= \alpha\delta - \gamma\beta \end{aligned}$$

(a) Show that Cdet is F-multilinear in the rows and columns of the matrix.
(b) Show that Cdet is *not* left B-multilinear in the rows of the matrix.
(c) Give an example showing that Cdet is not multiplicative.
(d) Find a matrix $A \in \mathrm{M}_2(\mathbb{H})$ that is invertible (i.e., having a two-sided inverse) but has $\mathrm{Cdet}(A) = 0$. Then find such an A with the further property that its transpose has nonzero determinant but is not invertible.

[Moral: be careful with matrix rings over noncommutative rings! For more on quaternionic determinants, including the Dieudonné determinant, see Aslaksen [Asl96].]

This chapter is licensed under the terms of the Creative Commons Attribution-NonCommercial 4.0 International License (http://creativecommons.org/licenses/by-nc/4.0/), which permits any noncommercial use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license and indicate if changes were made.

The images or other third party material in this chapter are included in the chapter's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the chapter's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.

Chapter 3

Involutions

In this chapter, we define the standard involution on a quaternion algebra. In this way, we characterize division quaternion algebras as noncommutative division rings equipped with a standard involution.

3.1 ▹ Conjugation

The quaternion conjugation map (2.4.6) defined on the Hamiltonians $\mathbb{H}$ arises naturally from the notion of real and pure (imaginary) parts, as defined by Hamilton. This involution has a natural generalization to a quaternion algebra $B = (a, b \mid F)$ over a field F with $\operatorname{char} F \neq 2$: we define

$$\begin{aligned} \overline{}: B &\to B \\ \alpha = t + xi + yj + zij &\mapsto \overline{\alpha} = t - (xi + yj + zij) \end{aligned}$$

Multiplying out, we then verify that

$$\alpha\overline{\alpha} = \overline{\alpha}\alpha = t^2 - ax^2 - by^2 + abz^2 \in F.$$

The way in which the cross terms cancel, because the basis elements i, j, k skew commute, is a calculation that never fails to enchant!

But this definition seems to depend on a basis: it is not intrinsically defined. What properties characterize it? Is it unique? We are looking for a good definition of conjugation $\overline{} : B \to B$ on an F-algebra B: we will call such a map a *standard involution*.

The involutions we consider should have basic linearity properties: they are F-linear (with $\overline{1} = 1$, so they act as the identity on F) and have order 2 as an F-linear map. An involution should also respect the multiplication structure on B, but we should not require that it be an F-algebra isomorphism: instead, like the inverse map (or transpose map) reverses order of multiplication, we ask that $\overline{\alpha\beta} = \overline{\beta}\,\overline{\alpha}$ for all $\alpha \in B$. Finally, we want the standard involution to give rise to a trace and norm (a measure of size), which is to say, we want $\alpha + \overline{\alpha} \in F$ and $\alpha\overline{\alpha} = \overline{\alpha}\alpha \in F$ for all $\alpha \in B$. The precise definition is given in Definition 3.2.1, and the defining properties are rigid: if an algebra B has a standard involution, then it is necessarily unique (Corollary 3.4.4).

© The Author(s) 2021

J. Voight, *Quaternion Algebras*, Graduate Texts in Mathematics 288,
https://doi.org/10.1007/978-3-030-56694-4_3

The existence of a standard involution on B implies that every element of B satisfies a quadratic equation: by direct substitution, we see that $\alpha \in B$ is a root of the polynomial $x^2 - tx + n \in F[x]$ where $t := \alpha + \overline{\alpha}$ and $n := \alpha\overline{\alpha} = \overline{\alpha}\alpha$, since then

$$\alpha^2 - (\alpha + \overline{\alpha})\alpha + \alpha\overline{\alpha} = 0$$

identically. Accordingly, we define the **reduced trace** $\operatorname{trd}\colon B \to F$ by $\operatorname{trd}(\alpha) = \alpha + \overline{\alpha}$ and **reduced norm** $\operatorname{nrd}\colon B \to F$ by $\operatorname{nrd}(\alpha) = \alpha\overline{\alpha}$. We observe that trd is F-linear and nrd is multiplicative on $B^\times$.

Motivated by this setting, we say that B has **degree** 2 if every element $\alpha \in B$ satisfies a (monic) polynomial in $F[x]$ of degree 2 and, to avoid trivialities, that $B \neq F$ (or equivalently, at least one element of B satisfies no polynomial of degree 1). The final result of this section is the following theorem (see Theorem 3.5.1).

Theorem 3.1.1. *Let B be a division F-algebra of degree 2 over a field F with* char $F \neq 2$. *Then either $B = K$ is a quadratic field extension of F or B is a division quaternion algebra over F.*

As a consequence, division quaternion algebras are characterized as noncommutative division algebras with a standard involution, when char $F \neq 2$.

3.2 Involutions

Throughout this chapter, let B be an F-algebra. For the moment, we allow F to be of arbitrary characteristic. We begin by defining involutions on B.

Definition 3.2.1. An **involution** $\overline{} : B \to B$ is an F-linear map which satisfies:

(i) $\overline{1} = 1$;
(ii) $\overline{\overline{\alpha}} = \alpha$ for all $\alpha \in B$; and
(iii) $\overline{\alpha\beta} = \overline{\beta}\,\overline{\alpha}$ for all $\alpha, \beta \in B$ (the map $\overline{}$ is an **anti-automorphism**).

3.2.2. We define the **opposite algebra** of B by letting $B^{\mathrm{op}} = B$ as F-vector spaces but with multiplication $\alpha \cdot_{\mathrm{op}} \beta = \beta \cdot \alpha$ for $\alpha, \beta \in B$.

One can then equivalently define an involution to be an F-algebra isomorphism $B \xrightarrow{\sim} B^{\mathrm{op}}$ whose underlying F-linear map has order at most 2.

Remark 3.2.3. What we have defined to be an involution is known in other contexts as an **involution of the first kind**. An **involution of the second kind** is a map which acts nontrivially when restricted to F, and hence is not F-linear; although these involutions are interesting in other contexts, they will not figure in our discussion (and anyway one can consider such an algebra over the fixed field of the involution).

Definition 3.2.4. An involution $\overline{}$ is **standard** if $\alpha\overline{\alpha} \in F$ for all $\alpha \in B$.

Remark 3.2.5. Standard involutions go by many other names. The terminology *standard* is employed because conjugation on a quaternion algebra is the "standard" example of such an involution. Other authors call the standard involution the *main involution*

for quaternion algebras, but then find situations where the "main" involution is not standard by our definition. The standard involution is also called *conjugation* on B, but this can be confused with conjugation by an element in $B^\times$. We will see in Corollary 3.4.4 that a standard involution is unique, so it is also called the *canonical involution*; however, there are other circumstances where involutions can be defined canonically that are not standard (like the map induced by $g \mapsto g^{-1}$ on the group ring $F[G]$).

3.2.6. If $\overline{}$ is a standard involution, so that $\alpha\overline{\alpha} \in F$ for all $\alpha \in B$, then

$$(\alpha + 1)\overline{(\alpha + 1)} = (\alpha + 1)(\overline{\alpha} + 1) = \alpha\overline{\alpha} + \alpha + \overline{\alpha} + 1 \in F$$

and hence $\alpha + \overline{\alpha} \in F$ for all $\alpha \in B$ as well; it then also follows that $\alpha\overline{\alpha} = \overline{\alpha}\alpha$, since

$$(\alpha + \overline{\alpha})\alpha = \alpha(\alpha + \overline{\alpha}).$$

Example 3.2.7. The identity map is a standard involution on $B = F$ as an F-algebra. The $\mathbb{R}$-algebra $\mathbb{C}$ has a standard involution, namely, complex conjugation.

Example 3.2.8. The **adjugate** map

$$A = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \mapsto A^\dagger = \begin{pmatrix} d & -b \\ -c & a \end{pmatrix}$$

is a standard involution on $M_2(F)$ since $AA^\dagger = A^\dagger A = ad - bc = \det A \in F$.

Matrix transpose is an involution on $M_n(F)$ but is a standard involution (if and) only if $n = 1$.

3.2.9. Suppose $\operatorname{char} F \neq 2$ and let $B = (a, b \mid F)$. Then the map

$$\begin{gathered} \overline{} : B \to B \\ \alpha = t + xi + yj + zij \mapsto \overline{\alpha} = t - xi - yj - zij \end{gathered}$$

defines a standard involution on B and $\overline{\alpha} = 2t - \alpha$. The map is F-linear with $\overline{1} = 1$ and $\overline{\overline{\alpha}} = \alpha$, so properties (i) and (ii) hold. By F-linearity, it is enough to check property (iii) on a basis (Exercise 3.1), and we verify for instance that

$$\overline{ij} = \overline{ij} = -ij = ji = (-j)(-i) = \overline{j}\,\overline{i}$$

(see Exercise 3.3). Finally, the involution is standard because

$$(t + xi + yj + zij)(t - xi - yj - zij) = t^2 - ax^2 - by^2 + abz^2 \in F. \tag{3.2.10}$$

Remark 3.2.11. Algebras with involution play an important role in analysis, in particular Banach algebras with involution and C^*-algebras (generally of infinite dimension). A good reference is the text by Dixmier [Dix77] (or the more introductory book by Conway [Con2012]).

3.3 Reduced trace and reduced norm

Let $\overline{}: B \to B$ be a standard involution on B. We define the **reduced trace** on B by

$$\begin{aligned} \operatorname{trd}\colon B &\to F \\ \alpha &\mapsto \alpha + \overline{\alpha} \end{aligned} \tag{3.3.1}$$

and similarly the **reduced norm**

$$\begin{aligned} \operatorname{nrd}\colon B &\to F \\ \alpha &\mapsto \alpha\overline{\alpha}. \end{aligned} \tag{3.3.2}$$

Example 3.3.3. For $B = \mathrm{M}_2(F)$, equipped with the adjugate map as a standard involution as in Example 3.2.8, the reduced trace is the usual matrix trace and the reduced norm is the determinant.

3.3.4. The reduced trace trd is an F-linear map, since this is true for the standard involution:

$$\operatorname{trd}(\alpha+\beta) = (\alpha+\beta) + \overline{(\alpha+\beta)} = (\alpha+\overline{\alpha}) + (\beta+\overline{\beta}) = \operatorname{trd}(\alpha) + \operatorname{trd}(\beta)$$

for $\alpha, \beta \in B$. The reduced norm nrd is multiplicative, since

$$\operatorname{nrd}(\alpha\beta) = (\alpha\beta)\overline{(\alpha\beta)} = \alpha\beta\overline{\beta}\,\overline{\alpha} = \alpha\operatorname{nrd}(\beta)\overline{\alpha} = \operatorname{nrd}(\alpha)\operatorname{nrd}(\beta)$$

for all $\alpha, \beta \in B$.

It will be convenient to write

$$\begin{aligned} B^0 &:= \{\alpha \in B : \operatorname{trd}(\alpha) = 0\} \\ B^1 &:= \{\alpha \in B^\times : \operatorname{nrd}(\alpha) = 1\} \end{aligned} \tag{3.3.5}$$

for the F-subspace $B^0 \subseteq B$ of elements of reduced trace 0 and for the subgroup $B^1 \leq B^\times$ of elements of reduced norm 1. We observe that $B^1 \trianglelefteq B^\times$ is normal, by multiplicativity, indeed we have an exact sequence of groups

$$1 \to B^1 \to B^\times \xrightarrow{\operatorname{nrd}} F^\times$$

(noting that the reduced norm map need not be surjective).

Lemma 3.3.6. *If B is not the zero ring, then $\alpha \in B$ is a unit (has a two-sided inverse) if and only if* $\operatorname{nrd}(\alpha) \neq 0$.

Proof. Exercise 3.5. ∎

Lemma 3.3.7. *For all $\alpha, \beta \in B$, we have* $\operatorname{trd}(\beta\alpha) = \operatorname{trd}(\alpha\beta)$.

Proof. We have

$$\operatorname{trd}(\alpha\overline{\beta}) = \operatorname{trd}(\alpha(\operatorname{trd}(\beta) - \beta)) = \operatorname{trd}(\alpha)\operatorname{trd}(\beta) - \operatorname{trd}(\alpha\beta)$$

and so

$$\operatorname{trd}(\alpha\overline{\beta}) = \operatorname{trd}(\overline{\alpha\overline{\beta}}) = \operatorname{trd}(\beta\overline{\alpha}) = \operatorname{trd}(\alpha)\operatorname{trd}(\beta) - \operatorname{trd}(\beta\alpha)$$

therefore $\operatorname{trd}(\alpha\beta) = \operatorname{trd}(\beta\alpha)$. □

Remark 3.3.8. The maps trd and nrd are called *reduced* for the following reason.

Let A be a finite-dimensional F-algebra, and consider the left regular representation $\lambda\colon A \hookrightarrow \operatorname{End}_F(A)$ given by left multiplication in A (cf. Proposition 2.3.1, but over F). We then have a (left) trace map $\operatorname{Tr}\colon A \to F$ and (left) norm map $\operatorname{Nm}\colon A \to F$ given by mapping $\alpha \in B$ to the trace and determinant of the endomorphism $\lambda_\alpha \in \operatorname{End}_F(A)$.

When $A = \mathrm{M}_2(F)$, a direct calculation (Exercise 3.13) reveals that

$$\operatorname{Tr}(\alpha) = 2\operatorname{trd}(\alpha) = 2\operatorname{tr}(\alpha)$$

(algebra trace, reduced trace, and matrix trace, respectively; there is no difference between left and right), and

$$\operatorname{Nm}(\alpha) = \operatorname{nrd}(\alpha)^2 = \det(\alpha)^2$$

for all $\alpha \in A$, whence the name *reduced*. (To preview the language of chapter 7, this calculation can be efficiently summarized: as a left A-module, A is the sum of two simple A-modules—acting on the columns of a matrix—and the reduced trace and reduced norm represent 'half' of this action.)

3.3.9. Since

$$\alpha^2 - (\alpha + \overline{\alpha})\alpha + \alpha\overline{\alpha} = 0 \tag{3.3.10}$$

identically we see that $\alpha \in B$ is a root of the polynomial

$$x^2 - \operatorname{trd}(\alpha)x + \operatorname{nrd}(\alpha) \in F[x] \tag{3.3.11}$$

which we call the **reduced characteristic polynomial** of α. The fact that α satisfies its reduced characteristic polynomial is the **reduced Cayley–Hamilton theorem** for an algebra with standard involution. When $\alpha \notin F$, the reduced characteristic polynomial of α is its minimal polynomial, since if α satisfies a polynomial of degree 1 then $\alpha \in F$.

3.4 Uniqueness and degree

Definition 3.4.1. An F-algebra K with $\dim_F K = 2$ is called a **quadratic** algebra.

Lemma 3.4.2. *Let K be a quadratic F-algebra. Then K is commutative and has a unique standard involution.*

Proof. Let $\alpha \in K \smallsetminus F$. Then $K = F \oplus F\alpha = F[\alpha]$, so in particular K is commutative. Then $\alpha^2 = t\alpha - n$ for unique $t, n \in F$, since $1, \alpha$ is a basis for K.

If $\overline{} : K \to K$ is any standard involution, then from (3.3.10) and uniqueness we conclude $t = \alpha + \overline{\alpha}$ (and $n = \alpha\overline{\alpha}$), and so any involution must have $\overline{\alpha} = t - \alpha$. On the other hand, there is a unique standard involution $\overline{x}\colon B \to B$ with $\overline{\alpha} = t - \alpha$: the verification is straightforward (see Exercise 3.2). □

Example 3.4.3. The reduced trace and norm on a quadratic algebra are precisely the usual algebra trace and norm. If $\operatorname{char} F \neq 2$ and $K \supseteq F$ is a quadratic field extension of F, then the standard involution is just the nontrivial element of $\operatorname{Gal}(K \mid F)$.

Corollary 3.4.4. *If B has a standard involution, then this involution is unique.*

Proof. For any $\alpha \in B \smallsetminus F$, we have from (3.3.10) that $\dim_F F[\alpha] = 2$, so the restriction of the standard involution to $F[\alpha]$ is unique. Therefore the standard involution on B is itself unique. □

We have seen that the equation (3.3.10), implying that if B has a standard involution then every $\alpha \in B$ satisfies a quadratic equation, has figured prominently in the above proofs. To further clarify the relationship between these two notions, we make the following definition.

Definition 3.4.5. The **degree** of B is the smallest $m \in \mathbb{Z}_{\geq 0}$ such that every element $\alpha \in B$ satisfies a monic polynomial $f(x) \in F[x]$ of degree m, if such an integer exists; otherwise, we say B has degree ∞.

3.4.6. If B has finite dimension $n = \dim_F B < \infty$, then every element of B satisfies a polynomial of degree at most n: if $\alpha \in B$ then the elements $1, \alpha, \dots, \alpha^n$ are linearly dependent over F. Consequently, every finite-dimensional F-algebra has a (well-defined) integer degree, at most n.

Example 3.4.7. By convention, we interpret Definition 3.4.5 as defining the degree of the zero ring to be 0 (since $1 = 0$, the element 0 satisfies the monic polynomial $0x$)—whatever!

If B has degree 1, then $B = F$. If B has a standard involution, then either $B = F$ or B has degree 2 by (3.3.11).

3.5 Quaternion algebras

We are now ready to characterize division algebras of degree 2 when $\operatorname{char} F \neq 2$. (For the case $\operatorname{char} F = 2$, see Chapter 6.)

Theorem 3.5.1. *Suppose* $\operatorname{char} F \neq 2$ *and let B be a division F-algebra. Then B has degree at most 2 if and only if one of the following hold:*

(i) *$B = F$;*

(ii) *$B = K$ is a quadratic field extension of F; or*

(iii) *B is a division quaternion algebra over F.*

Proof. From Example 3.4.7, we may suppose that $B \neq F$ and B has degree 2.

Let $i \in B \smallsetminus F$. Then $F[i] = K$ is a (commutative) quadratic F-subalgebra of the division ring B, so $K = F(i)$ is a field. If $K = B$, we are done. Completing the square (since $\operatorname{char} F \neq 2$), we may suppose that $i^2 = a \in F^\times$.

Let $\phi \colon B \to B$ be the map given by conjugation by i, i.e., $\phi(\alpha) = i^{-1}\alpha i$. Then ϕ is a K-linear endomorphism of B, thought of as a (left) K-vector space, and ϕ^2 is the identity on B. Therefore ϕ is diagonalizable, and we may decompose $B = B^+ \oplus B^-$ into eigenspaces for ϕ: explicitly, we can always write

$$\alpha = \frac{\alpha + \phi(\alpha)}{2} + \frac{\alpha - \phi(\alpha)}{2} \in B^+ \oplus B^-.$$

We now prove $\dim_K B^+ = 1$. Let $\alpha \in B^+$. Then $L = F(\alpha, i)$ is a field. Since $\operatorname{char} F \neq 2$, and L is a compositum of quadratic extensions of F, the primitive element theorem implies that $L = F(\beta)$ for some $\beta \in L$. But by hypothesis β satisfies a quadratic equation so $\dim_F L = 2$ and hence $L = K$. (For an alternative direct proof of this claim, see Exercise 3.10.)

If $B = B^+ = K$, we are done. So suppose $B^- \neq \{0\}$. We will prove that $\dim_K B^- = 1$. If $0 \neq j \in B^-$ then $i^{-1}ji = -j$, so $i = -j^{-1}ij$ and hence all elements of B^- conjugate i to $-i$. Thus if $0 \neq j_1, j_2 \in B^-$ then $j_1 j_2$ centralizes i and $j_1 j_2 \in B^+ = K$. Thus any two nonzero elements of B^- are K-multiples of each other.

Finally, let $j \in B^- \smallsetminus \{0\}$; then $B = B^+ \oplus B^- = K \oplus Kj$ so B has F-basis $1, i, j, ij$ and $ji = -ij$. We claim that $\operatorname{trd}(j) = 0$: indeed, both j and $i^{-1}ji = -j$ satisfy the same reduced characteristic (or minimal) polynomial of degree 2, so $\operatorname{trd}(j) = \operatorname{trd}(-j) = -\operatorname{trd}(j)$ so $\operatorname{trd}(j) = 0$. Thus $j^2 = b \in F^\times$, and B is a quaternion algebra by definition. □

Remark 3.5.2. We need not assume in Theorem 3.5.1 that B is finite-dimensional; somehow, it is a consequence, and every division algebra over F (with $\operatorname{char} F \neq 2$) of degree ≤ 2 is finite-dimensional.

There are algebras of arbitary (finite or infinite) dimension over F of degree 2: see Exercise 3.15. Also, a boolean ring (see Exercise 3.12) has degree 2 as an $\mathbb{F}_2$-algebra, and there are such rings of arbitrary dimension over $\mathbb{F}_2$. Such algebras are quite far from being division rings, of course.

Remark 3.5.3. The proof of Theorem 3.5.1 has quite a bit of history, discussed by van Praag [vPr2002] (along with several proofs). See Lam [Lam2005, Theorem III.5.1] for a parallel proof of Theorem 3.5.1. Moore [Moore35, Theorem 14.4] in 1915 studied algebra of matrices over skew fields and in particular the role of involutions, and gives an elementary proof of this theorem (with the assumption $\operatorname{char} F \neq 2$). Dieudonné [Die48, Die53] gave another proof that relies on structure theory for finite-dimensional division algebras.

Corollary 3.5.4. *Let B be a division F-algebra with* $\operatorname{char} F \neq 2$. *Then B has degree at most* 2 *if and only if B has a standard involution.*

Proof. In each of the cases (i)–(iii), B has a standard involution; and conversely if B has a standard involution, then B has degree at most 2 (Example 3.4.7). □

Remark 3.5.5. The statement of Corollary 3.5.4 holds more generally—even if B is not necessarily a division ring—as follows. Let B be an F-algebra with $\operatorname{char} F \neq 2$. Then B has a standard involution if and only if B has degree at most 2 [Voi2011b]. However, this is no longer true in characteristic 2 (Exercise 3.12).

Corollary 3.5.6. *Let B be a division F-algebra with* $\operatorname{char} F \neq 2$. *Then the following are equivalent:*

(i) *B is a quaternion algebra;*
(ii) *B is noncommutative and has degree* 2*; and*
(iii) *B is central and has degree* 2*.*

Definition 3.5.7. An F-algebra B is **algebraic** if every $\alpha \in B$ is algebraic over F (i.e., α satisfies a polynomial with coefficients in F).

If B has finite degree (such as when $\dim_F B = n < \infty$), then B is algebraic.

Corollary 3.5.8. (Frobenius). *Let B be an algebraic division algebra over $\mathbb{R}$. Then either $B = \mathbb{R}$ or $B \simeq \mathbb{C}$ or $B \simeq \mathbb{H}$ as $\mathbb{R}$-algebras.*

Proof. If $\alpha \in B \smallsetminus \mathbb{R}$ then $\mathbb{R}(\alpha) \simeq \mathbb{C}$, so α satisfies a polynomial of degree 2. Thus if $B \neq \mathbb{R}$ then B has degree 2 and either $B \simeq \mathbb{C}$ or B is a division quaternion algebra over $\mathbb{R}$, and hence $B \simeq \mathbb{H}$ by Exercise 2.4(c). □

Example 3.5.9. Division algebras over $\mathbb{R}$ of infinite dimension abound. Transcendental field extensions of $\mathbb{R}$, such as the function field $\mathbb{R}(x)$ or the Laurent series field $\mathbb{R}((x))$, are examples of infinite-dimensional division algebras over $\mathbb{R}$. Also, the free algebra in two (noncommuting) variables is a subring of a division ring B (its "noncommutative ring of fractions") with center $\mathbb{R}$ and of infinite dimension over $\mathbb{R}$.

Remark 3.5.10. The theorem of Frobenius (Corollary 3.5.8) extends directly to fields F akin to $\mathbb{R}$, as follows. A field is **formally real** if -1 cannot be expressed in F as a sum of squares and **real closed** if F is formally real and has no formally real proper algebraic extension. The real numbers $\mathbb{R}$ and the field of all real algebraic numbers are real closed. A real closed field has characteristic zero, is totally ordered, and contains a square root of each nonnegative element; the field obtained from F by adjoining a root of the irreducible polynomial $x^2 + 1$ is algebraically closed. For these statements, see Rajwade [Raj93, Chapter 15]. Every finite-dimensional division algebra over a real closed field F is either F or $K = F(\sqrt{-1})$ or $B = (-1, -1 \mid F)$.

Remark 3.5.11. Algebras of dimension 3, sitting somehow between quadratic extensions and quaternion algebras, can be characterized in a similar way. If B is an $\mathbb{R}$-algebra of dimension 3, then either B is commutative or B has a standard involution and is isomorphic to the subring of upper triangular matrices in $\mathrm{M}_2(\mathbb{R})$. A similar statement holds for free R-algebras of rank 3 over a (commutative) domain R; see Levin [Lev2013].

Exercises

Throughout these exercises, let F be a field.

▹ 1. Let B be an F-algebra and let $\overline{}: B \to B$ be an F-linear map with $\overline{1} = 1$. Show that $\overline{}$ is an involution if and only if (ii)–(iii) in Definition 3.2.1 hold for a basis of B (as an F-vector space).

▹ 2. Let $K = F[\alpha]$ be a quadratic F-algebra, with $\alpha^2 = t\alpha - n$ for (unique) $t, n \in F$. Extending linearly, show that there is a unique standard involution $\overline{}: K \to K$ with the property that $\overline{\alpha} = t - \alpha$, and show that

$$\operatorname{trd}(x + y\alpha) = 2x + ty$$
$$\operatorname{nrd}(x + y\alpha) = x^2 + txy + ny^2$$

for all $x + y\alpha \in F[\alpha]$.

▹ 3. Verify that the map $\overline{}$ in Example 3.2.9 is a standard involution.

4. Determine the standard involution on $K = F \times F$ (with $F \hookrightarrow K$ under the diagonal map).

▹ 5. Let B be an F-algebra with a standard involution. Show that $0 \neq \alpha \in B$ is a left zerodivisor if and only if α is a right zerodivisor if and only if $\operatorname{nrd}(\alpha) = 0$. In particular, if B is not the zero ring, then $\alpha \in B$ is (left and right) invertible if and only if $\operatorname{nrd}(\alpha) \neq 0$.

6. Suppose $\operatorname{char} F \neq 2$, let B be a division quaternion algebra over F, and let $K_1, K_2 \subseteq B$ be subfields with $K_1 \cap K_2 = F$. Show that the F-subalgebra of B generated by K_1 and K_2 is equal to B. Conclude that if $1, \alpha, \beta \in B$ are F-linearly independent, then $1, \alpha, \beta, \alpha\beta$ are an F-basis for B. *[Hint: use the involution.]* By way of counterexample, show that these results need not hold for $B = \mathrm{M}_2(F)$.

7. Show that $B = \mathrm{M}_n(F)$ has a standard involution if and only if $n \leq 2$.

8. Let G be a finite group. Show that the F-linear map induced by $g \mapsto g^{-1}$ for $g \in G$ is an involution on the group ring $F[G] = \bigoplus_{g \in G} Fg$. Determine necessary and sufficient conditions for this map to be a standard involution.

9. Let B be an F-algebra with a standard involution $\overline{} : B \to B$. In this exercise, we examine when $\overline{}$ is the identity map.

   (a) Show that if $\operatorname{char} F \neq 2$, then $x \in B$ satisfies $\overline{x} = x$ if and only $x \in F$.

   (b) Suppose that $\dim_F B < \infty$. Show that the identity map is a standard involution on B if and only if (i) $B = F$ or (ii) $\operatorname{char} F = 2$ and B is a quotient of the commutative ring $F[x_1, \dots, x_n]/(x_1^2 - a_1, \dots, x_n^2 - a_n)$ with $a_i \in F$.

10. Let $K \supseteq F$ be a field which has degree m as an F-algebra in the sense of Definition 3.4.5. Suppose that $\operatorname{char} F \nmid m$. Show that $[K : F] = m$, i.e., K has degree m in the usual sense. (What happens when $\operatorname{char} F \mid m$?)

11. Let B be an F-algebra with standard involution. Suppose that $\phi: B \xrightarrow{\sim} B$ is an F-algebra automorphism. Show for $\alpha \in B$ that $\overline{\phi(\alpha)} = \phi(\overline{\alpha})$, and therefore

that $\overline{\operatorname{trd}(\phi(\alpha))} = \operatorname{trd}(\alpha)$ and $\operatorname{nrd}(\phi(\alpha)) = \operatorname{nrd}(\alpha)$. *[Hint: consider the map* $\alpha \mapsto \phi^{-1}(\overline{\phi(\alpha)})$.]

12. In this exercise, we explore further the relationship between algebras of degree 2 and those with standard involutions (Remark 3.5.5).

    (a) Suppose char $F \neq 2$ and let B be a finite-dimensional F-algebra. Show that B has a standard involution if and only if $\deg_F B \leq 2$.
    (b) Let $F = \mathbb{F}_2$ and let B be a Boolean ring, a ring such that $x^2 = x$ for all $x \in B$. (Verify that $2 = 0$ in B, so B is an $\mathbb{F}_2$-algebra.) Prove that B does not have a standard involution unless $B = \mathbb{F}_2$ or $B = \mathbb{F}_2 \times \mathbb{F}_2$, but nevertheless any Boolean ring has degree at most 2.

▹ 13. Let $B = \mathrm{M}_n(F)$, and consider the map $\lambda\colon B \hookrightarrow \operatorname{End}_F(B)$ by $\alpha \mapsto \lambda_\alpha$ defined by left-multiplication in B. Show that for all $\alpha \in \mathrm{M}_n(F)$, the characteristic polynomial of λ_α is the nth power of the usual characteristic polynomial of α. Conclude when $n = 2$ that $\operatorname{tr}(\alpha) = 2\operatorname{trd}(A)$ and $\det(\alpha) = \operatorname{nrd}(\alpha)^2$.

14. Considering a slightly different take on the previous exercise: let B be a quaternion algebra over F. Show that the characteristic polynomial of left multiplication by $\alpha \in B$ is equal to that of right multiplication and is the square of the reduced characteristic polynomial. *[Hint: if a direct approach is too cumbersome, consider applying the previous exercise and the left regular representation as in 2.3.8.]*

15. Let V be an F-vector space and let $t : V \to F$ be an F-linear map. Let $B = F \oplus V$ and define the binary operation $x \cdot y = t(x)y$ for $x, y \in V$. Show that $\cdot$ induces a multiplication on B, and that the map $x \mapsto \overline{x} = t(x) - x$ for $x \in V$ induces a standard involution on B. [Such an algebra is called an **exceptional algebra** [GrLu2009, Voi2011b].] Conclude that there exists a central F-algebra B with a standard involution in any dimension $n = \dim_F B \geq 3$.

▹ 16. In this exercise, we mimic the proof of Theorem 3.5.1 to prove that a quaternion algebra over a finite field of odd cardinality is not a division ring, a special case of Wedderburn's little theorem: a finite division ring is a field.

    Assume for purposes of contradiction that B is a division quaternion algebra over $F = \mathbb{F}_q$ with q odd.

    (a) Let $i \in B \smallsetminus F$. Show that the centralizer $C_{B^\times}(i) = \{\alpha \in B^\times : i\alpha = \alpha i\}$ of i in $B^\times$ satisfies $C_{B^\times}(i) = F(i)^\times$.
    (b) Conclude that any noncentral conjugacy class in $B^\times$ has order $q^2 + 1$.
    (c) Derive a contradiction from the class equation $q^4 - 1 = q - 1 + m(q^2 + 1)$ (where $m \in \mathbb{Z}$).

    [For the case q even, see Exercise 6.16; for fun, the eager reader may wish to prove Weddernburn's little theorem for $F = \mathbb{F}_2$ directly.]

17. Derive Euler's identity (1.1.6) that the product of the sum of four squares is again the sum of four squares as follows. Let $F = \mathbb{Q}(x_1, \ldots, x_4, y_1, \ldots, y_4)$ be a function field over $\mathbb{Q}$ in 8 variables and consider the quaternion algebra $(-1, -1 \mid F)$. Show (by an explicit universal formula) that if R is any commutative ring and $x, y \in R$ are the sum of four squares in R, then xy is the sum of four squares in R.
18. Suppose char $F \neq 2$. For an F-algebra B, let

$$V(B) = \{\alpha \in B \smallsetminus F : \alpha^2 \in F\} \cup \{0\}.$$

Let B be a division ring. Show that $V(B)$ is a vector space (closed under addition) if and only if $B = F$ or $B = K$ is a quadratic field extension of F or B is a quaternion algebra over F.
19. Let B be an F-algebra with F-basis $e_1, e_2, \ldots, e_n$. Let $\overline{} : B \to B$ be an involution. Show that $\overline{}$ is standard if and only if

$$e_i\overline{e_i} \in F \text{ and } (e_i + e_j)\overline{(e_i + e_j)} \in F \text{ for all } i, j = 1, \ldots, n.$$

This chapter is licensed under the terms of the Creative Commons Attribution-NonCommercial 4.0 International License (http://creativecommons.org/licenses/by-nc/4.0/), which permits any noncommercial use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license and indicate if changes were made.

The images or other third party material in this chapter are included in the chapter's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the chapter's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.

Chapter 4

Quadratic forms

Quaternion algebras, as algebras equipped with a standard involution, are intrinsically related to quadratic forms. We develop this connection in the next two chapters.

4.1 ▹ Reduced norm as quadratic form

Let F be a field with $\operatorname{char} F \neq 2$ and let $B = (a, b \mid F)$ be a quaternion algebra over F. We have seen (3.2.9) that B has a unique standard involution and consequently a reduced norm map, with

$$\operatorname{nrd}(t + xi + yj + zij) = t^2 - ax^2 - by^2 + abz^2 \tag{4.1.1}$$

for $t, x, y, z \in F$. The reduced norm therefore defines a **quadratic form**, a homogeneous polynomial of degree 2 in $F[t, x, y, z]$ (thought of as a function of the coefficients of an element respect to the basis $1, i, j, k$). It should come as no surprise, then, that the structure of the quaternion algebra B is related to properties of the quadratic form nrd.

Let $Q\colon V \to F$ be a quadratic form. Then Q can be diagonalized by a change of variables: there is a basis $e_1, \dots, e_n$ of V such that

$$Q(x_1 e_1 + \dots + x_n e_n) = Q(x_1, \dots, x_n) = a_1 x_1^2 + \dots + a_n x_n^2$$

with $a_i \in F$. We define the **discriminant** of Q to be the (well-defined) product $\operatorname{disc}(Q) := a_1 \cdots a_n / 2^n \in F/F^{\times 2}$. (The factor 2^n is for consistency with more general notions; it is harmless if a bit annoying.) We say that a quadratic form is **nondegenerate** if its discriminant is nonzero. The reduced norm quadratic form (4.1.1) is already diagonal in the basis $1, i, j, k$, and it is nondegenerate because $a, b \neq 0$.

A **similarity** from Q to another quadratic form $Q'\colon V' \to F$ is a pair (f, u) where $f\colon V \xrightarrow{\sim} V'$ is an F-linear isomorphism and $u \in F^\times$ satisfy $Q'(f(x)) = uQ(x)$ for all $x \in V$. An **isometry** is a similarity with $u = 1$. The **orthogonal group** of Q is the group of self-isometries of Q, i.e.,

$$\mathrm{O}(Q)(F) := \{f \in \operatorname{Aut}_F(V) : Q(f(x)) = Q(x) \text{ for all } x \in V\}.$$

© The Author(s) 2021

J. Voight, *Quaternion Algebras*, Graduate Texts in Mathematics 288,
https://doi.org/10.1007/978-3-030-56694-4_4

An isometry $f \in \mathrm{O}(Q)(F)$ is **special** if $\det f$, and the **special orthogonal group** of Q is the group of special isometries of Q.

More generally, we have seen that any algebra with a standard involution has a quadratic form nrd. We say that the standard involution is **nondegenerate** whenever the quadratic form nrd is so. Generalizing Theorem 3.1.1, we prove the following (see Main Theorem 4.4.1 for the proof).

Main Theorem 4.1.2. *Let B be an F-algebra. Then B has a nondegenerate standard involution if and only if one of the following holds:*

(i) $B = F$;
(ii) $B = K$ *has* $\dim_F K = 2$ *and either* $K \simeq F \times F$ *or* K *is a field; or*
(iii) B *is a quaternion algebra over* F.

This theorem gives another way of characterizing quaternion algebras: they are noncommutative algebras with a nondegenerate standard involution.

In Section 2.4, we saw that the unit Hamiltonians $\mathbb{H}^1$ act on the pure Hamiltonians $\mathbb{H}^0$ (Section 2.4) by *rotations*: the standard Euclidean quadratic form (sum of squares) is preserved by conjugation. This generalizes in a natural way to an arbitrary field, and so we can understand the group of linear transformations that preserve a ternary (or quaternary) form in terms of the unit group of a quaternion algebra B (Proposition 4.5.10): there is an exact sequence

$$1 \to F^\times \to B^\times \to \mathrm{SO}(\mathrm{nrd}|_{B^0})(F) \to 1$$

where $\mathrm{SO}(Q)(F)$ is the group of special (or oriented) isometries of the quadratic form Q.

4.2 Basic definitions

In this section, we summarize basic definitions and notation for quadratic forms over fields. The "Bible for all quadratic form practitioners" (according to the MathSciNet review by K. Szymiczek) is the book by Lam [Lam2005]; in particular, Lam gives a very readable account of the relationship between quadratic forms and quaternion algebras over F when $\operatorname{char} F \neq 2$ [Lam2005, Sections III.1–III.2] and many other topics in the algebraic theory of quadratic forms. Also recommended are the books by Cassels [Cas78], O'Meara [O'Me73], and Scharlau [Scha85], as well as the book by Grove [Grov2002], who treats quadratic forms from a geometric point of view in terms of the orthogonal group. For reference and further inspiration, see also the hugely influential book by Eichler [Eic53].

Let F be a field. (For now, we allow $\operatorname{char} F$ to be arbitrary.)

Definition 4.2.1. A **quadratic form** Q is a map $Q\colon V \to F$ on an F-vector space V satisfying:

(i) $Q(ax) = a^2Q(x)$ for all $a \in F$ and $x \in V$; and
(ii) The map $T\colon V \times V \to F$ defined by

$$T(x, y) = Q(x + y) - Q(x) - Q(y)$$

is F-bilinear.

We call the pair (V, Q) a **quadratic space** and T the **associated bilinear form**.

We will often abbreviate a quadratic space (V, Q) by simply V. If Q is a quadratic form then the associated bilinear form T is **symmetric**, satisfying $T(x, y) = T(y, x)$ for all $x, y \in V$; in particular, $T(x, x) = 2Q(x)$ for all $x \in V$, so when $\operatorname{char} F \neq 2$ we recover the quadratic form from the symmetric bilinear form.

For the remainder of this section, let $Q\colon V \to F$ be a quadratic form with associated bilinear form T.

4.2.2. Suppose $\dim_F V = n < \infty$. Let $e_1, \dots, e_n$ be a basis for V, giving an isomorphism $V \simeq F^n$. Then Q can be written

$$Q(x_1e_1 + \dots + x_ne_n) = \sum_i Q(e_i)x_i^2 + \sum_{i<j} T(e_i, e_j)x_ix_j \in F[x_1, \dots, x_n]$$

as a homogeneous polynomial of degree 2.

The **Gram matrix** of Q in the basis e_i is the (symmetric) matrix

$$[T] := (T(e_i, e_j))_{i,j} \in \mathrm{M}_n(F).$$

We then have $T(x, y) = x^{\mathsf{t}}[T]y$ for $x, y \in V \simeq F^n$ as column vectors. Under a change of basis $A \in \mathrm{GL}_n(F)$ with $e_i' = Ae_i$, the Gram matrix $[T]'$ in the basis e_i' has

$$[T]' = A^{\mathsf{t}}[T]A. \tag{4.2.3}$$

Definition 4.2.4. A **similarity** of quadratic forms from $Q\colon V \to F$ to $Q'\colon V' \to F$ is a pair (f, u) where $f\colon V \xrightarrow{\sim} V'$ is an F-linear isomorphism and $u \in F^\times$ satisfy $Q'(f(x)) = uQ(x)$ for all $x \in V$, i.e., such that the diagram

$$\begin{array}{ccc} V & \xrightarrow{\;Q\;} & F \\ \wr\downarrow f & & \wr\downarrow u \\ V' & \xrightarrow{\;Q'\;} & F \end{array} \tag{4.2.5}$$

commutes. In a similarity (f, u), the scalar u is called the **similitude factor** of the similarity. An **isometry** of quadratic forms (or **isomorphism** of quadratic spaces) is a similarity with similitude factor $u = 1$; we write in this case $Q \simeq Q'$.

Definition 4.2.6. The **general orthogonal group** (or **similarity group**) of the quadratic form Q is the group of self-similarities of Q under composition

$$\mathrm{GO}(Q)(F) := \{(f, u) \in \mathrm{Aut}_F(V) \times F^\times : Q(f(x)) = uQ(x) \text{ for all } x \in V\};$$

the **orthogonal group** of Q is the group of **self-isometries** of Q, i.e.,

$$\mathrm{O}(Q)(F) := \{f \in \mathrm{Aut}_F(V) : Q(f(x)) = Q(x) \text{ for all } x \in V\}.$$

Remark 4.2.7. A similarity allows isomorphisms of the target F (as a one-dimensional F-vector space). The notion of *isometry* comes from the connection with measuring lengths, when working with the usual Euclidean norm form on a vector space over $\mathbb{R}$: *similarity* allows these lengths to scale uniformly (e.g., similar triangles).

There is a canonical exact sequence

$$\begin{aligned} 1 \to \mathrm{O}(Q)(F) \to \mathrm{GO}(Q)(F) &\to F^\times \\ (f, u) &\mapsto u \end{aligned} \tag{4.2.8}$$

realizing $\mathrm{O}(Q)(F) \leq \mathrm{GO}(Q)(F)$ as the subgroup of self-similarities with similitude factor $u = 1$.

4.2.9. Returning to 4.2.2, suppose $\dim_F V = n < \infty$ and $\operatorname{char} F \neq 2$. Then one can understand the orthogonal group of Q quite concretely in matrix terms as follows. Choose a basis $e_1, \dots, e_n$ for V and let $[T]$ be the Gram matrix of Q with respect to this basis, so that $2Q(x) = x^{\mathsf{t}}[T]x$ for all $x \in V \simeq F^n$. Then $\operatorname{Aut}_F(V) \simeq \mathrm{GL}_n(F)$ and $A \in \mathrm{GL}_n(F)$ belongs to $\mathrm{O}(Q)$ if and only if

$$(Ax)^{\mathsf{t}}[T](Ax) = x^{\mathsf{t}}(A^{\mathsf{t}}[T]A)x = x^{\mathsf{t}}[T]x$$

for all $x \in V$, and therefore

$$\mathrm{O}(Q)(F) = \{A \in \mathrm{GL}_n(F) : A^{\mathsf{t}}[T]A = [T]\} \tag{4.2.10}$$

and

$$\mathrm{GO}(Q)(F) = \{(A, u) \in \mathrm{GL}_n(F) \times F^\times : A^{\mathsf{t}}[T]A = u[T]\} \tag{4.2.11}$$

From now on, let $Q \colon V \to F$ be a quadratic form and let $T : V \times V \to F$ be the symmetric bilinear form associated to Q.

Definition 4.2.12. Let $x, y \in V$. We say that x is **orthogonal** to y (with respect to Q) if $T(x, y) = 0$.

Since T is symmetric, x is orthogonal to y if and only if y is orthogonal to x for $x, y \in V$, and so we simply say x, y are orthogonal. If $S \subseteq V$ is a subset, we write

$$S^\perp := \{x \in V : T(v, x) = 0 \text{ for all } v \in S\}$$

for the subspace of V which is orthogonal to (the span of) S.

4.2.13. Let B be an algebra over F with a standard involution. Then $\operatorname{nrd} \colon B \to F$ is a quadratic form on B. Indeed, $\operatorname{nrd}(a\alpha) = a^2\alpha$ for all $\alpha \in B$, and the map T given by

$$T(\alpha, \beta) = (\alpha + \beta)\overline{(\alpha + \beta)} - \alpha\overline{\alpha} - \beta\overline{\beta} = \alpha\overline{\beta} + \beta\overline{\alpha} = \alpha\overline{\beta} + \overline{\alpha\overline{\beta}} = \operatorname{trd}(\alpha\overline{\beta}) \tag{4.2.14}$$

for $\alpha, \beta \in B$ is bilinear, and

$$T(\alpha, \beta) = \operatorname{trd}(\alpha\overline{\beta}) = \operatorname{trd}(\alpha(\operatorname{trd}(\beta) - \beta)) = \operatorname{trd}(\alpha)\operatorname{trd}(\beta) - \operatorname{trd}(\alpha\beta). \tag{4.2.15}$$

So $\alpha, \beta \in B$ are orthogonal with respect to nrd if and only if

$$\operatorname{trd}(\alpha\overline{\beta}) = \alpha\overline{\beta} + \beta\overline{\alpha} = 0$$

if and only if

$$\operatorname{trd}(\alpha\beta) = \operatorname{trd}(\alpha)\operatorname{trd}(\beta).$$

Thus 1 and $\alpha \in B$ are orthogonal if and only if $\operatorname{trd}(\alpha) = 0$ if and only if $\alpha^2 = -\operatorname{nrd}(\alpha)$. Moreover, rearranging (4.2.14),

$$\alpha\beta + \beta\alpha = \operatorname{trd}(\beta)\alpha + \operatorname{trd}(\alpha)\beta - T(\alpha, \beta). \tag{4.2.16}$$

In particular, if $1, \alpha, \beta \in B$ are linearly independent over F, then by (4.2.16) they are pairwise orthogonal if and only if $\beta\alpha = -\alpha\beta$.

In this way, we see that the multiplication law in B is governed in a fundamental way by the reduced norm quadratic form.

Definition 4.2.17. Let $Q \colon V \to F$ be a quadratic form. We say that Q **represents** an element $a \in F$ if there exists $x \in V$ such that $Q(x) = a$. A quadratic form is **universal** if it represents every element of F.

Definition 4.2.18. A quadratic form Q (or a quadratic space V) is **isotropic** if Q represents 0 nontrivially (there exists $0 \neq x \in V$ such that $Q(x) = 0$) and otherwise Q is **anisotropic**.

Remark 4.2.19. The terminology *isotropic* is as least as old as Eichler [Eic53, p. 3], and perhaps it goes back to Witt. The word can be used to mean “having properties that are identical in all directions”, and so the motivation for this language possibly comes from physics: the second fundamental form associated to a parametrized surface $z = f(x, y)$ in $\mathbb{R}^3$ is a quadratic form, and (roughly speaking) this quadratic form defines the curvature at a given point. In this sense, if the quadratic form vanishes, then the curvature is zero, and things look the same in all directions.

4.2.20. Let $Q' \colon V' \to F$ be another quadratic form. We define the **orthogonal direct sum**

$$\begin{gathered} Q \boxplus Q' \colon V \oplus V' \to F \\ (Q \boxplus Q')(x + x') = Q(x) + Q(x') \end{gathered}$$

where $x \in V$ and $x' \in V'$; the associated bilinear form $T \boxplus T'$ has

$$(T \boxplus T')(x + x', y + y') = T(x, y) + T(x', y')$$

for all $x, y \in V$ and $x', y' \in V'$. By definition, under the natural inclusion of $V, V' \subseteq V \oplus V'$, we have $V' \subseteq V^\perp$ (and $V \subseteq (V')^\perp$).

4.2.21. For $a \in F$, we write $\langle a \rangle$ for the quadratic form ax^2 on F. More generally, for $a_1, \dots, a_n \in F$, we write

$$\langle a_1 \rangle \boxplus \dots \boxplus \langle a_n \rangle := \langle a_1, \dots, a_n \rangle$$

for the quadratic form on F^n defined by $Q(x_1, \dots, x_n) = a_1 x_1^2 + \dots + a_n x_n^2$.

To conclude this introduction, we state an important result due originally to Witt which governs the decomposition of quadratic spaces into orthogonal sums up to isometry.

Theorem 4.2.22. *Let $V \simeq V'$ be isometric quadratic spaces with orthogonal decompositions $V \simeq W_1 \boxplus W_2$ and $V' \simeq W_1' \boxplus W_2'$.*

(a) *If $W_1 \simeq W_1'$, then $W_2 \simeq W_2'$.*
(b) *If $g : W_1 \xrightarrow{\sim} W_1'$ is an isometry, then there exists an isometry $f : V \xrightarrow{\sim} V'$ such that $f|_{W_1} = g$ and $f(W_2) = W_2'$.*

Proof. The proof is requested in Exercise 4.16. For a proof and the equivalence between Witt cancellation (part (a)) and Witt extension (part (b)), see Lam [Lam2005, Proof of Theorem I.4.2, p. 14], Scharlau [Scha85, Theorem 1.5.3], or O'Meara [O'Me73, Theorem 42:17]. ◻

Theorem 4.2.22(a) is called **Witt cancellation** and 4.2.22(b) is called **Witt extension**.

4.3 Discriminants, nondegeneracy

For the remainder of this chapter, we suppose that $\operatorname{char} F \neq 2$. (We take up the case $\operatorname{char} F = 2$ in section 6.3.) Throughout, let $Q \colon V \to F$ be a quadratic form with $\dim_F V = n < \infty$ and associated symmetric bilinear form T.

The following result (proven by induction) is a standard application of Gram–Schmidt orthogonalization (Exercise 4.1); working with a quadratic form as a polynomial, this procedure can be thought of as iteratively *completing the square*.

Lemma 4.3.1. *There exists a basis of V such that $Q \simeq \langle a_1, \dots, a_n\rangle$ with $a_i \in F$.*

A form presented with a basis as in Lemma 4.3.1 is called **normalized** (or **diagonal**). For a diagonal quadratic form Q, the associated Gram matrix $[T]$ is diagonal with entries $2a_1, \dots, 2a_n$.

4.3.2. The determinant $\det([T])$ of a Gram matrix for Q depends on a choice of basis for V, but by (4.2.3), a change of basis matrix $A \in \mathrm{GL}_n(F)$ operates on $[T]$ by $A^{\mathsf{t}}[T]A$, and $\det(A^{\mathsf{t}}[T]A) = \det(A)^2 \det([T])$, so we obtain a well-defined element $\det(T) \in F/F^{\times 2}$ independent of the choice of basis.

Definition 4.3.3. The **discriminant** of Q is

$$\operatorname{disc}(Q) := 2^{-n} \det(T) \in F/F^{\times 2}.$$

The **signed discriminant** of Q is

$$\operatorname{sgndisc}(Q) := (-1)^{n(n-1)/2} \operatorname{disc}(Q) \in F/F^{\times 2}.$$

When it will cause no confusion, we will represent the class of the discriminant in $F/F^{\times 2}$ simply by a representative element in F.

Remark 4.3.4. The extra factor 2^{-n} is harmless since $\operatorname{char} F \neq 2$, and it allows us to naturally cancel certain factors 2 that appear whether we are in even or odd dimension—it will be essential when we consider the case $\operatorname{char} F = 2$ (see 6.3.1). The distinction between even and odd dimensional quadratic spaces is not arbitrary: indeed, this distinction is pervasive, even down to the classification of semisimple Lie algebras.

Example 4.3.5. We have $\operatorname{disc}(\langle a_1, \dots, a_n \rangle) = a_1 \cdots a_n$ for $a_i \in F$.

Definition 4.3.6. The bilinear form $T : V \times V \to F$ is **nondegenerate** if for all $x \in V \smallsetminus \{0\}$, the linear functional $T_x \colon V \to F$ defined by $T_x(y) = T(x, y)$ is nonzero, i.e., there exists $y \in V$ such that $T(x, y) \neq 0$. We say that Q (or V) is **nondegenerate** if the associated bilinear form T is nondegenerate.

4.3.7. The bilinear form T induces a map

$$\begin{aligned} V &\to \operatorname{Hom}(V, F) \\ x &\mapsto (y \mapsto T(x, y)) \end{aligned}$$

and T is nondegenerate if and only if this map is injective (and hence an isomorphism) if and only if $\det(T) \neq 0$. Put another way, Q is nondegenerate if and only if $\operatorname{disc}(Q) \neq 0$, and so a diagonal form $\langle a_1, \dots, a_n \rangle$ is nondegenerate if and only if $a_i \neq 0$ for all i.

Example 4.3.8. Let $B = (a, b \mid F)$ be a quaternion algebra. Then by 3.2.9, the quadratic form $\operatorname{nrd} \colon B \to F$ is normalized with respect to the basis $1, i, j, ij$. Indeed,

$$\operatorname{nrd} \simeq \langle 1, -a, -b, ab \rangle.$$

We have $\operatorname{disc}(\operatorname{nrd}) = (ab)^2 \neq 0$, so nrd is nondegenerate.

If B is an F-algebra with a standard involution, then the reduced norm defines a quadratic form on B, and we say that the standard involution is **nondegenerate** if nrd is nondegenerate.

4.3.9. One can often restrict to the case where a quadratic form Q is nondegenerate by *splitting off the radical*, as follows. We define the **radical** of Q to be

$$\operatorname{rad}(Q) := V^{\perp} = \{x \in V : T(x, y) = 0 \text{ for all } y \in V\}.$$

The radical $\operatorname{rad}(Q) \subseteq V$ is a subspace, so completing a basis of $\operatorname{rad}(Q)$ to V we can write (noncanonically) $V = \operatorname{rad}(Q) \boxplus W$, as the direct sum is an orthogonal direct sum by definition of the radical. In this decomposition, $Q|_{\operatorname{rad}(Q)}$ is identically zero and $Q|_W$ is nondegenerate.

4.4 Nondegenerate standard involutions

In this section, we follow Theorem 3.5.1 with a characterization of quaternion algebras beyond division algebras.

Main Theorem 4.4.1. *Suppose* char $F \neq 2$ *and let B be an F-algebra. Then B has a nondegenerate standard involution if and only if one of the following holds:*

(i) *$B = F$;*
(ii) *$B = K$ is a quadratic F-algebra and either $K \simeq F \times F$ or K is a field; or*
(iii) *B is a quaternion algebra over F.*

Case (ii) in Main Theorem 4.4.1 is equivalent to requiring that K be a quadratic F-algebra that is **reduced** (has no nonzero nilpotent elements).

Remark 4.4.2. By Exercise 3.15, there exist F-algebras with standard involution having arbitrary dimension, so it is remarkable that the additional requirement that the standard involution be nondegenerate gives such a tidy result.

Proof of Main Theorem 4.4.1. If $B = F$, then the standard involution is the identity and nrd is nondegenerate. If $\dim_F K = 2$, then after completing the square we may write $K \simeq F[x]/(x^2 - a)$ and in the basis $1, x$ we find nrd $\simeq \langle 1, a\rangle$. By Example 4.3.5, nrd is nondegenerate if and only if $a \in F^\times$ if and only if K is a quadratic field extension of F or $K \simeq F \times F$.

Suppose that $\dim_F B > 2$. Let $1, i, j$ be a part of a normalized basis for B with respect to the quadratic form nrd. Then $T(1, i) = \operatorname{trd}(i) = 0$, so $i^2 = a \in F^\times$, since nrd is nondegenerate. Note in particular that $\overline{i} = -i$. Similarly $j^2 = b \in F^\times$, and by (4.2.16) we have $\operatorname{trd}(ij) = ij + ji = 0$. We have $T(1, ij) = \operatorname{trd}(ij) = 0$, and $T(ij, i) = \operatorname{trd}(\overline{i}(ij)) = -a\operatorname{trd}(j) = 0$ and similarly $T(ij, j) = 0$, hence $ij \in \{1, i, j\}^\perp$. If $ij = 0$ then $i(ij) = aj = 0$ so $j = 0$, a contradiction. Since nrd is nondegenerate, it follows then that the set $1, i, j, ij$ is linearly independent.

Therefore, the subalgebra A of B generated by i, j satisfies $A \simeq (a, b \mid F)$, and if $\dim_F B = 4$ we are done. So let $k \in A^\perp$; then $\operatorname{trd}(k) = 0$ and $k^2 = c \in F^\times$. Thus $k \in B^\times$, with $k^{-1} = c^{-1}k$. By 4.2.13 we have $k\alpha = \overline{\alpha}k$ for any $\alpha \in A$ since $\overline{k} = -k$. But then

$$k(ij) = (\overline{ij})k = \overline{j}\,\overline{i}k = \overline{j}ki = k(ji). \tag{4.4.3}$$

But $k \in B^\times$ so $ij = ji = -ij$, and this is a contradiction. □

Main Theorem 4.4.1 has the following corollaries.

Corollary 4.4.4. *Let B be an F-algebra with* char $F \neq 2$*. Then B is a quaternion algebra if and only if B is noncommutative and has a nondegenerate standard involution.*

Proof. Immediate. □

Corollary 4.4.5. *Let B have a nondegenerate standard involution, and suppose that $K \subseteq B$ is a commutative F-subalgebra such that the restriction of the standard involution is nondegenerate. Then* $\dim_F K \leq 2$*. Moreover, if $K \neq F$, then the centralizer of $K^\times$ in $B^\times$ is $K^\times$.*

Proof. The first statement is immediate; the second follows by considering the algebra generated by the centralizer. □

Remark 4.4.6. Algebras with involutions come from quadratic forms, and the results of this chapter are just one special case of a much more general theory. More precisely, there is a natural bijection between the set of isomorphism classes of finite-dimensional simple F-algebras equipped with an F-linear involution and the set of similarity classes of nondegenerate quadratic forms on finite-dimensional F-vector spaces. More generally, for involutions that act nontrivially on the base field, one looks at Hermitian forms. Consequently, there are three broad types of involutions on central simple algebras, depending on the associated quadratic or Hermitian form: orthogonal, symplectic, and unitary. Accordingly, algebras with involutions can be classified by the invariants of the associated form. This connection is the subject of the tome by Knus–Merkurjev–Rost–Tignol [KMRT98]. In this way the theory of quadratic forms belongs to the theory of algebras with involution, which in turn is a part of the theory of linear algebraic groups, as expounded by Weil [Weil60]: see the survey by Tignol [Tig98] for an overview and further references.

4.5 Special orthogonal groups

In this section, we revisit the original motivation of Hamilton (Section 2.4) in a more general context, relating quaternions to the orthogonal group of a quadratic form. We retain our running hypothesis that $\operatorname{char} F \neq 2$ and $Q\colon V \to F$ is a nondegenerate quadratic form with $\dim_F V = n < \infty$.

Definition 4.5.1. An isometry $f \in \mathrm{O}(Q)(F)$ is **special** (or **proper**) if $\det f = 1$. The **special orthogonal group** of Q is the group of special isometries of Q:

$$\mathrm{SO}(Q)(F) := \{f \in \mathrm{O}(Q)(F) : \det(f) = 1\}.$$

The condition "$\det f = 1$" is well-defined, independent of the choice of F-basis of V; having chosen a basis of V so that $\mathrm{O}(Q)(F) \leq \mathrm{GL}_n(F)$, we have $\mathrm{SO}(Q)(F) = \mathrm{O}(Q)(F) \cap \mathrm{SL}_n(F)$.

4.5.2. Suppose that $V = F^n$ and let $f \in \mathrm{O}(Q)$ be a self-isometry of Q, represented in the standard basis by $A \in \mathrm{GL}_n(F)$. Taking determinants in (4.2.10) we conclude that $\det(A)^2 = 1$ so $\det(A) = \pm 1$. The determinant is surjective (see Exercise 4.15), so we have an exact sequence

$$1 \to \mathrm{SO}(Q)(F) \to \mathrm{O}(Q)(F) \xrightarrow{\det} \{\pm 1\} \to 1.$$

If n is *odd*, then either f or $-f$ is special, so the sequence splits and

$$\mathrm{O}(Q)(F) \simeq \{\pm 1\} \times \mathrm{SO}(Q)(F). \tag{4.5.3}$$

4.5.4. Similarly, if $(f, u) \in \mathrm{GO}(Q)(F)$ then from (4.2.11) we get $u^{-n}\det(f)^2 = 1$. If $n = 2m$ is even, then $u^{-m}\det(f) = \pm 1$, and we define the **general special orthogonal group** (or **special similarity group**) of Q to be

$$\mathrm{GSO}(Q)(F) := \{(f, u) \in \mathrm{GO}(Q)(F) : u^{-m}\det(f) = 1\}$$

giving an exact sequence

$$1 \to \mathrm{GSO}(Q)(F) \to \mathrm{GO}(Q)(F) \to \{\pm 1\} \to 1.$$

If n is odd, we have little choice other than to define $\mathrm{GSO}(Q)(F) := \mathrm{GO}(Q)(F)$.

Example 4.5.5. If $V = \mathbb{R}^n$ and Q is the usual Euclidean norm on V, then

$$\mathrm{O}(Q)(\mathbb{R}) = \mathrm{O}(n) = \{A \in \mathrm{GL}_n(\mathbb{R}) : AA^{\mathsf{t}} = 1\}$$

is the group of linear maps preserving length (but not necessarily orientation), whereas $\mathrm{SO}(Q)(\mathbb{R})$ is the usual group of rotations of V (preserving orientation). Similarly, $\mathrm{GSO}(Q)(\mathbb{R})$ consists of orientation-preserving similarities, preserving orientation but allowing a constant scaling.

In particular, if $n = 2$ then $\mathrm{O}(2) := \mathrm{O}(Q)(\mathbb{R})$ contains

$$\mathrm{SO}(2) := \mathrm{SO}(Q)(\mathbb{R}) = \left\{ \begin{pmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{pmatrix} : \theta \in \mathbb{R} \right\} \simeq \mathbb{R}/(2\pi\mathbb{Z}) \simeq \mathbb{S}^1$$

(the circle group) with index 2, with a reflection in any line through the origin representing a nontrivial coset of $\mathrm{SO}(2) \le \mathrm{O}(2)$.

4.5.6. More generally, we may define reflections in $\mathrm{O}(Q)(F)$ as follows. For $x \in V$ anisotropic (so $Q(x) \neq 0$), we define the **reflection** in x to be

$$\begin{aligned} \tau_x &: V \to V \\ \tau_x(v) &= v - \frac{T(v,x)}{Q(x)} x. \end{aligned}$$

We have $\tau_x(x) = x - 2x = -x$, and

$$\begin{aligned} Q(\tau_x(v)) &= Q(v) + Q\left(-\frac{T(v,x)}{Q(x)} x\right) + T\left(v, -\frac{T(v,x)}{Q(x)} x\right) \\ &= Q(v) + \frac{T(v,x)^2}{Q(x)^2} Q(x) - \frac{T(v,x)}{Q(x)} T(v,x) = Q(v) \end{aligned}$$

so $\tau_x \in \mathrm{O}(Q)(F) \smallsetminus \mathrm{SO}(Q)(F)$.

By a classical theorem of Cartan and Dieudonné, the orthogonal group is generated by reflections.

Theorem 4.5.7. (Cartan–Dieudonné). *Let (V, Q) be a nondegenerate quadratic space with $\dim_F V = n$. Then every isometry $f \in \mathrm{O}(Q)(F)$ is a product of at most n reflections.*

Proof. See Lam [Lam2005, §I.7], O'Meara [O'Me73, §43B], or Scharlau [Scha85, Theorem 1.5.4]. The proof is by induction on n, carefully recording the effect of a reflection in an anisotropic vector. □

Since reflections have determinant -1, an isometry f is special (and $f \in \mathrm{SO}(Q)(F)$) if and only if it is the product of an *even* number of reflections.

4.5.8. Now let B be a quaternion algebra over F, and recall (3.3.5) that we have defined

$$B^0 := \{v \in B : \operatorname{trd}(B) = 0\}.$$

Then there is a (left) action

$$\begin{aligned} B^\times \circlearrowright B^0 &\to B^0 \\ \alpha \cdot v &= \alpha v \alpha^{-1}. \end{aligned} \tag{4.5.9}$$

since $\operatorname{trd}(\alpha v \alpha^{-1}) = \operatorname{trd}(v) = 0$. Moreover, $B^\times$ acts on V by isometries with respect to the quadratic form $Q = \operatorname{nrd}|_{B^0} : V \to F$, since $\operatorname{nrd}(\alpha v \alpha^{-1}) = \operatorname{nrd}(v)$ for all $\alpha \in B$ and $v \in V$.

Proposition 4.5.10. *Let B be a quaternion algebra over F. Then the action* (4.5.9) *induces an exact sequence*

$$1 \to F^\times \to B^\times \to \mathrm{SO}(\operatorname{nrd}|_{B^0})(F) \to 1. \tag{4.5.11}$$

If further $\operatorname{nrd}(B^\times) = F^{\times 2}$, *then*

$$1 \to \{\pm 1\} \to B^1 \to \mathrm{SO}(\operatorname{nrd}|_{B^0})(F) \to 1,$$

where $B^1 := \{\alpha \in B : \operatorname{nrd}(\alpha) = 1\}$.

Proof. Let $Q = \operatorname{nrd}|_{B^0}$. We saw in 4.5.8 that the action of $B^\times$ is by isometries, so lands in $\mathrm{O}(Q)(F)$. By the Cartan–Dieudonné theorem (Theorem 4.5.7, the weak version of Exercise 4.17 suffices), every isometry is the product of reflections, and by determinants an isometry is special if and only if it is the product of an even number of reflections. A reflection in $x \in V = B^0$ with $Q(x) = \operatorname{nrd}(x) \neq 0$ is of the form

$$\begin{aligned} \tau_x(v) &= v - \frac{T(v,x)}{Q(x)} x = v - \frac{\operatorname{trd}(v\overline{x})}{\operatorname{nrd}(x)} x \\ &= v - (v\overline{x} + x\overline{v})\overline{x}^{-1} = -x\overline{v}\,\overline{x}^{-1} = x\overline{v}x^{-1}, \end{aligned} \tag{4.5.12}$$

the final equality from $\overline{x} = -x$ as $x \in B^0$. The product of two such reflections is thus of the form $v \mapsto \alpha v \alpha^{-1}$ with $\alpha \in B^\times$. Therefore $B^\times$ acts by special isometries, and every special isometry so arises: the map $B^\times \to \mathrm{O}(Q)(F)$ surjects onto $\mathrm{SO}(Q)(F)$. The kernel of the action is given by those $\alpha \in B^\times$ with $\alpha v \alpha^{-1} = v$ for all $v \in B^0$, i.e., $\alpha \in Z(B^\times) = F^\times$.

The second statement follows directly by writing $B^\times = B^1 F^\times$. □

Example 4.5.13. If $B \simeq \mathrm{M}_2(F)$, then nrd = det, so $\det^0 \simeq \langle 1, -1, -1 \rangle$ and (4.5.11) yields the isomorphism $\mathrm{PGL}_2(F) \simeq \mathrm{SO}(\langle 1, -1, -1 \rangle)(F)$.

Example 4.5.14. If $F = \mathbb{R}$ and $B = \mathbb{H}$, then $\det(\mathbb{H}) = \mathbb{R}_{>0} = \mathbb{R}^{\times 2}$, and the second exact sequence is Hamilton's (Section 2.4).

To conclude, we pass from three variables to four variables.

4.5.15. In Exercise 2.17, we showed that there is an exact sequence

$$1 \to \{\pm 1\} \to \mathbb{H}^1 \times \mathbb{H}^1 \to \mathrm{SO}(4) \to 1$$

with $\mathbb{H}^1 \times \mathbb{H}^1$ acting on $\mathbb{H} \simeq \mathbb{R}^4$ by $v \mapsto \alpha v \beta^{-1} = \alpha v \overline{\beta}$ for $\alpha, \beta \in \mathbb{H}^1$.

More generally, let B be a quaternion algebra over F. Then there is a left action of $B^\times \times B^\times$ on B:

$$\begin{aligned} B^\times \times B^\times \circlearrowright B &\to B \\ (\alpha, \beta) \cdot v &= \alpha v \beta^{-1}. \end{aligned} \tag{4.5.16}$$

This action is by similarities, since if $a = \mathrm{nrd}(\alpha)$ and $b = \mathrm{nrd}(\beta)$, then

$$\mathrm{nrd}(\alpha v \beta^{-1}) = \mathrm{nrd}(\alpha)\,\mathrm{nrd}(v)\,\mathrm{nrd}(\beta^{-1}) = \frac{a}{b}\,\mathrm{nrd}(v)$$

for all $v \in V$, with similitude factor $u = a/b$. In particular, if $\mathrm{nrd}(\alpha) = \mathrm{nrd}(\beta)$, then the action is by isometries.

Proposition 4.5.17. *With notation as in 4.5.15, the left action (4.5.16) induces exact sequences*

$$\begin{aligned} 1 \to F^\times \to B^\times \times B^\times &\to \mathrm{GSO}(\mathrm{nrd})(F) \to 1 \\ a \mapsto (a, a)& \end{aligned} \tag{4.5.18}$$

and

$$1 \to F^\times \to \{(\alpha, \beta) \in B^\times \times B^\times : \mathrm{nrd}(\alpha) = \mathrm{nrd}(\beta)\} \to \mathrm{SO}(\mathrm{nrd})(F) \to 1.$$

If further $\mathrm{nrd}(B^\times) = F^{\times 2}$*, then the sequence*

$$1 \to \{\pm 1\} \to B^1 \times B^1 \to \mathrm{SO}(\mathrm{nrd})(F) \to 1$$

is exact.

Proof. For the first statement, we first show that the kernel of the action is the diagonally embedded $F^\times$. Suppose that $\alpha v \beta^{-1} = v$ for all $v \in B$; taking $v = 1$ shows $\beta = \alpha$, and then we conclude that $\alpha v = v \alpha$ for all $v \in B$ so $\alpha \in Z(B) = F$.

Next, the map $B^\times \times B^\times \to \mathrm{GSO}(\mathrm{nrd})(F)$ is surjective. If $f \in \mathrm{GSO}(\mathrm{nrd})(F)$ then $\mathrm{nrd}(f(x)) = u\,\mathrm{nrd}(x)$ for all $x \in B$, so in particular $u \in \mathrm{nrd}(B^\times)$. Every such similitude factor occurs, since the similitude factor of $(\alpha, 1)$ is $\mathrm{nrd}(\alpha)$. So it suffices to show that the map

$$\{(\alpha, \beta) \in B^\times \times B^\times : \mathrm{nrd}(\alpha) = \mathrm{nrd}(\beta)\} \to \mathrm{SO}(\mathrm{nrd})(F)$$

is surjective. We again appeal to the Cartan–Dieudonné theorem; by the same computation as in (4.5.12), we calculate that a reflection in $x \in B^\times$ is of the form

$$\tau_x(v) = -x \overline{v}\, \overline{x}^{-1}.$$

The product of two reflections for $x, y \in B^\times$ is thus of the form

$$v \mapsto -y\overline{(-x\overline{v}\,\overline{x}^{-1})}\,\overline{y}^{-1} = (yx^{-1})v\overline{(yx^{-1})}^{-1} = \alpha v \beta^{-1}$$

where $\alpha = yx^{-1}$ and $\beta = \overline{\alpha}$, and in particular the action is by special similarities. We conclude that (4.5.18) and the second sequence are both exact.

The final statement again follows by writing $B^\times = B^1 F^\times$, and seeing the kernel as $F^\times \cap B^1 = \{\pm 1\}$. □

Example 4.5.19. When $B = \mathrm{M}_2(F)$, then $\mathrm{nrd}(B^\times) = \det(\mathrm{GL}_2(F)) = F^\times$, giving the exact sequence

$$1 \to \mathrm{GL}_1(F) \to \mathrm{GL}_2(F) \times \mathrm{GL}_2(F) \to \mathrm{GSO}(\det)(F) \to 1.$$

Exercises

Let F be a field with $\operatorname{char} F \neq 2$.

▷ 1. Give an algorithmic proof that every finite-dimensional quadratic space has a normalized basis (Lemma 4.3.1).

2. Let $F = \mathbb{R}$ and let

$$V = \left\{ (a_n)_n : a_n \in \mathbb{R} \text{ for all } n \geq 0 \text{ and } \sum_{n=0}^{\infty} a_n^2 \text{ converges} \right\}.$$

Show that V is an $\mathbb{R}$-vector space, and the map $Q\colon V \to \mathbb{R}$ by $Q((a_n)_n) = \sum_{n=0}^{\infty} a_n^2$ is a quadratic form, and so V is an example of an infinite-dimensional quadratic space. [This example generalizes to the context of *Hilbert spaces*.]

3. Let B be a quaternion algebra over F. Let $N : B \to F$ and $\Delta : B \to F$ be defined by $N(\alpha) = \mathrm{trd}(\alpha^2)$ and $\Delta(\alpha) = \mathrm{trd}(\alpha)^2 - 4\,\mathrm{nrd}(\alpha)$. Show that N, Δ are quadratic forms on B, describe their associated bilinear forms, and compute a normalized form (and basis) for each.

4. Generalize Exercise 2.15 as follows. Let $Q\colon V \to F$ be a quadratic form with $\dim_F V = n$ and let $\gamma \in \mathrm{O}(Q)$.

   (a) If n is odd and $\det \gamma = 1$, then γ has a nonzero fixed vector (and therefore restricts to the identity on a one-dimensional subspace of V).

   (b) If n is even and $\det \gamma = -1$, then γ has both eigenvalues -1 and 1.

5. Generalizing part of Exercise 4.3, let B be an F-algebra with a standard involution. Show that the **discriminant form**

$$\begin{aligned} \Delta\colon B &\to F \\ \Delta(\alpha) &= \mathrm{trd}(\alpha)^2 - 4\,\mathrm{nrd}(\alpha) \end{aligned}$$

is a quadratic form.

6. Let $Q\colon V \to F$ be a quadratic form with $\dim_F V < \infty$ and associated bilinear form T. The map

$$\begin{aligned} V &\to \mathrm{Hom}_F(V, F) \\ x &\mapsto (y \mapsto T(x, y)) \end{aligned}$$

is F-linear. Show that Q is nondegenerate if and only if this map is an isomorphism.

7. Write out the action (4.5.9) explicitly, as follows. Let $B = (a, b \mid F)$ and let $\alpha = t + xi + yj + zij$.

   (a) Show that the matrix of the action $v \mapsto \alpha v \alpha^{-1}$ in the F-basis $\beta = \{i, j, k\}$ for B^0 is $[\alpha]$ where $\operatorname{nrd}(\alpha)[\alpha]$ is equal to

   $$\begin{pmatrix} t^2 - ax^2 + by^2 - abz^2 & -2a(tz + xy) & -2(ty + axz) \\ 2b(tz - xy) & t^2 + ax^2 - by^2 - abz^2 & 2(tx - byz) \\ 2b(axz - ty) & 2a(tx + byz) & t^2 + ax^2 + by^2 + abz^2 \end{pmatrix}$$

   and $\operatorname{nrd}(\alpha) = t^2 - ax^2 - by^2 + abz^2$.

   (b) Let $Q = \operatorname{nrd}|_{B^0}$ and let T be the associated bilinear form. Show that the Gram matrix $[T]$ in the basis β is the diagonal matrix with entries $-2a, -2b, 2ab$. Then confirm by direct calculation that

   $$[\alpha] \in \mathrm{SO}(Q)(F) = \{A \in \mathrm{SL}_3(F) : A[T]A^{\mathsf{t}} = [T]\}.$$

8. In this exercise, we prove the **chain lemma**. Let $B := (a, b \mid F)$ be a quaternion algebra.

   (a) Show that if i' is orthogonal to $1, j$, then $(i')^2 = a' \in F^\times$ and i', j are standard generators for B, so $B \simeq (a', b \mid F)$.

   (b) Let $B' := (a', b' \mid F)$, and suppose that B is isomorphic to B'. Show that there exists $c \in F^\times$ such that

   $$B = \left(\frac{a, b}{F}\right) \simeq \left(\frac{c, b}{F}\right) \simeq \left(\frac{c, b'}{F}\right) \simeq \left(\frac{a', b'}{F}\right).$$

   [Hint: let $\phi\colon B' \xrightarrow{\sim} B$ be the isomorphism, and take an element orthogonal to $1, j, \phi(j')$.]

9. In this exercise, we develop some of the notions mentioned in Remark 3.3.8 in the context of quadratic forms.

   Let B be a finite-dimensional F-algebra (not necessarily a quaternion algebra), and let $\operatorname{Tr} : B \to F$ be the left algebra trace (the trace of the endomorphism given by left multiplication).

   (a) Show that the map $B \to F$ defined by $x \mapsto \operatorname{Tr}(x^2)$ is a quadratic form on B; this form is called the **(left) trace form** on B.

   (b) Compute the trace form of $A \times B$ and $A \otimes_F B$ in terms of the trace form of A and B.

   (c) Show that if $K \supseteq F$ is a inseparable field extension of finite degree, then the trace form on K (as an F-algebra) is identically zero. On the other hand, show that if K/F is a finite separable field extension (with char $F \neq 2$) then the trace form is nondegenerate.

   (d) Compute the trace form on $\mathbb{Q}(\sqrt{5})$ and $\mathbb{Q}(\alpha)$ where $\alpha = 2\cos(2\pi/7)$, so that $\alpha^3 + \alpha^2 - 2\alpha - 1 = 0$.

▹ 10. Let $Q: V \to F$ and $Q': V' \to F$ be quadratic forms over F with $\dim_F V = \dim_F V' = n < \infty$, and let T, T' be the associated bilinear forms. Suppose that there is a similarity $Q \sim Q'$ with similitude factor $u \in F^\times$. Show that $\det T' = u^n \det T \in F/F^{\times 2}$.

11. Let $Q: V \to F$ be a nondegenerate quadratic form with $\dim_F V = n < \infty$.

(a) A subspace $W \subseteq V$ is **totally isotropic** if $Q|_W = 0$ is identically zero. The **Witt index** $\nu(Q)$ of Q is the maximal dimension of a totally isotropic subspace. Show that if $\nu(Q) = m$ then $2m \leq n$.

(b) A **Pfister form** is a form in 2^m variables defined inductively by $\langle\langle a \rangle\rangle = \langle 1, -a \rangle$ and

$$\langle\langle a_1, \ldots, a_{m-1}, a_m \rangle\rangle = \langle\langle a_1, \ldots, a_{m-1} \rangle\rangle \boxplus -a_m \langle\langle a_1, \ldots, a_{m-1} \rangle\rangle.$$

Show that the reduced norm nrd on $\left(\dfrac{a, b}{F}\right)$ is the Pfister form $\langle\langle a, b \rangle\rangle$.

(c) The **hyperbolic plane** is the quadratic form $H: F^2 \to F$ with $H(x, y) = xy$. A quadratic form Q is a **hyperbolic plane** if $Q \simeq H$. A quadratic form Q is **totally hyperbolic** if $Q \simeq H \boxplus \ldots \boxplus H$ where H is a hyperbolic plane. Show that if Q is an isotropic Pfister form, then Q is totally hyperbolic.

(d) Suppose that Q is an isotropic Pfister form with $n \geq 4$. Let $W \subset V$ be a subspace of dimension $n - 1$. Show that $Q|_W$ is isotropic. [This gives another proof of Main Theorem 5.4.4 (iii) ⇒ (iv).]

12. (a) Let B be a quaternion algebra over F. Show that the reduced norm is the unique nonzero quadratic form Q on B that is **multiplicative**, i.e., $Q(\alpha\beta) = Q(\alpha)Q(\beta)$ for all $\alpha, \beta \in B$.

(b) Show that (a) does not necessarily hold more generally, for B an algebra with a standard involution. *[Hint: consider upper triangular matrices.]*

▹ 13. In this exercise, we pursue some geometric notions for readers some background in algebraic geometry (at the level of Hartshorne [Har77, Chapter 1]).
Let Q be nonzero quadratic form on V with $\dim_F V = n$. The vanishing locus of $Q(x) = 0$ defines a projective variety $X \subseteq \mathbb{P}(V) \simeq \mathbb{P}^n$ of degree 2 called a **quadric**. Show that the quadratic form Q is nondegenerate if and only if the projective variety X is nonsingular. [For this reason, a nondegenerate quadratic form is also synonymously called **nonsingular**.]

14. In this exercise, we work out from scratch Example 4.5.13: we translate the results on rotations in section 2.4 to $B = \mathrm{M}_2(\mathbb{R})$, but with respect to a different measure of 'length'.

Let

$$\mathrm{M}_2(\mathbb{R})^0 = \{v \in \mathrm{M}_2(\mathbb{R}) : \mathrm{tr}(v) = 0\} = \left\{\begin{pmatrix} x & y \\ z & -x \end{pmatrix} : x, y, z \in \mathbb{R}\right\}.$$

For $v \in \mathrm{M}_2(\mathbb{R})^0$, we have $\det(v) = -x^2 - yz$. Show that the group

$$\mathrm{M}_2(\mathbb{R})^1 = \mathrm{SL}_2(\mathbb{R}) = \{\alpha \in \mathrm{M}_2(\mathbb{R}) : \det(\alpha) = 1\}$$

acts linearly on $\mathrm{M}_2(\mathbb{R})^0$ by conjugation (the adjoint representation) preserving the determinant, giving rise to an exact sequence

$$1 \to \{\pm 1\} \to \mathrm{SL}_2(\mathbb{R}) \to \mathrm{SO}(\det) \to 1.$$

▹ 15. Let $Q: V \to F$ be a quadratic form with V finite-dimensional over F. Show that $\mathrm{SO}(Q) \leq \mathrm{O}(Q)$ is a (normal) subgroup of index 2. What can you say about $\mathrm{GSO}(Q) \leq \mathrm{GO}(Q)$?

▹ 16. In this exercise, we prove Theorem 4.2.22. Let $Q: V \to F$ be a quadratic form with $\dim_F V < \infty$ and let T be its associated bilinear form.

(a) Let $v \in V$ be anisotropic. Define the **reflection along** v by

$$\tau_v : V \to V$$
$$\tau_v(x) = x - \frac{T(v,x)}{Q(v)} v.$$

Observe that τ_v is F-linear, and then show that $\tau_v \in \mathrm{O}(V)$ with $\det \tau_v = -1$. *[Hint: extend v to a basis of the orthogonal complement of V.]* Why is τ_v called a reflection?

(b) If $x, y \in V$ are anisotropic with $Q(x) = Q(y)$, show that there exists $f \in \mathrm{O}(V)$ such that $f(x) = y$. *[Hint: reflect along either $v = x + y$ or $v = x - y$ as at least one is anisotropic, in the former case postcomposing with reflection along x.]*

(c) Let $Q': V' \to F$ be another quadratic form, and let $f: V \xrightarrow{\sim} V'$ be an isometry. For $W \subseteq V$, show that $f(W^{\perp}) = f(W)^{\perp}$.

(d) Prove Theorem 4.2.22(a). *[Hint: reduce to the case where* $\dim_F W_1 = \dim_F W_1' = 1$; *apply parts (b) and (c).]*

(e) Prove Theorem 4.2.22(b). *[Hint: compare the isometry $V \simeq V'$ with the isometry g.]*

▹ 17. Prove the following weakened version of the Cartan–Dieudonné theorem (Theorem 4.5.7): Let (V, Q) be a nondegenerate quadratic space with $\dim_F V = n$. Show that every isometry $f \in \mathrm{O}(Q)(F)$ is a product of at most $2n-1$ reflections. *[Hint: in the proof of Exercise 4.16(b), note that f can be taken to be a product of at most 2 reflections, and finish by induction.]*

This chapter is licensed under the terms of the Creative Commons Attribution-NonCommercial 4.0 International License (http://creativecommons.org/licenses/by-nc/4.0/), which permits any noncommercial use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license and indicate if changes were made.

The images or other third party material in this chapter are included in the chapter's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the chapter's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.

Chapter 5

Ternary quadratic forms and quaternion algebras

Continuing our treatment of quadratic forms, in this chapter we connect quaternion algebras to ternary quadratic forms.

5.1 ▹ Reduced norm as quadratic form

Let F be a field with $\operatorname{char} F \neq 2$ and let $B = (a, b \mid F)$ be a quaternion algebra over F. We saw in the previous chapter (4.1.1) that the reduced norm defines a quadratic form. But always have scalar norms $\operatorname{nrd}(t) = t^2$ for $t \in F$, so the form carries the same information when restricted to the space of pure quaternions

$$B^0 := \{\alpha \in B : \operatorname{trd}(\alpha) = 0\}$$

with basis i, j, ij. This quadratic form restricted to B^0 is

$$\operatorname{nrd}(xi + yj + zij) = -ax^2 - by^2 + abz^2$$

with discriminant $(-a)(-b)(ab) = (ab)^2$, so the trivial class in $F^\times/F^{\times 2}$.

We might now try to classify quaternion algebras over F up to isomorphism in terms of this quadratic form. Recall as in the previous chapters that for morphisms between quadratic forms, one allows either *isometries*, an invertible change of basis preserving the quadratic form, or *similarities*, which allow a rescaling of the quadratic form by a nonzero element of F. Our main result is as follows (Corollary 5.2.6).

Theorem 5.1.1. *The map* $B \mapsto \operatorname{nrd}|_{B^0}$ *induces a bijection:*

$$\left\{\begin{array}{c}\text{Quaternion algebras over } F\\ \text{up to isomorphism}\end{array}\right\} \leftrightarrow \left\{\begin{array}{c}\text{Ternary quadratic forms over } F\\ \text{with discriminant } 1 \in F^\times/F^{\times 2}\\ \text{up to isometry}\end{array}\right\}$$

$$\leftrightarrow \left\{\begin{array}{c}\text{Nondegenerate ternary}\\ \text{quadratic forms over } F\\ \text{up to similarity}\end{array}\right\}$$

© The Author(s) 2021

J. Voight, *Quaternion Algebras*, Graduate Texts in Mathematics 288,
https://doi.org/10.1007/978-3-030-56694-4_5

The map $B \mapsto \operatorname{nrd}|_{B^0}$ in Theorem 5.1.1 has inverse defined by the *even Clifford algebra* (see section 5.3). The similarity class of a nondegenerate ternary quadratic form cuts out a well-defined plane conic $C \subseteq \mathbb{P}^2$ over F, so one also has a bijection between isomorphism classes of quaternion algebras over F and isomorphism classes of conics over F. Finally, keeping track of an orientation allows one to fully upgrade this bijection to an equivalence of categories (Theorem 5.6.8).

The classification of quaternion algebras over F is now rephrased in terms of quadratic forms, and a more detailed description depends on the field F. In this vein, the most basic question we can ask about a quaternion algebra B is if it is isomorphic to the matrix ring $B \simeq \mathrm{M}_2(F)$: if so, we say that B is **split** over F. For example, every quaternion algebra over $\mathbb{C}$ (or an algebraically closed field) is split, and a quaternion algebra $(a, b \mid \mathbb{R})$ is split if and only if $a > 0$ or $b > 0$.

Ultimately, we will identify six equivalent ways (Main Theorem 5.4.4) to check if a quaternion algebra B is split; in light of Theorem 5.1.1, we isolate the following.

Proposition 5.1.2. *B is split if and only if the quadratic form $\operatorname{nrd}|_{B^0}$ represents 0 nontrivially.*

In later chapters, we will return to this classification problem, gradually increasing the "arithmetic complexity" of the field F.

5.2 Isomorphism classes of quaternion algebras

In Section 2.4, we found that the unit Hamiltonians act by conjugation on the pure quaternions $\mathbb{H}^0 \simeq \mathbb{R}^3$ as rotations, preserving the standard inner product. In this section, we return to this theme for a general quaternion algebra, and we characterize isomorphism classes of quaternion algebras in terms of isometry classes of ternary quadratic forms.

Throughout this chapter, let F be a field with $\operatorname{char} F \neq 2$, and let $B = (a, b \mid F)$ be a quaternion algebra over F.

Definition 5.2.1. $\alpha \in B$ is **scalar** if $\alpha \in F$ and **pure** if $\operatorname{trd}(\alpha) = 0$.

5.2.2. Recalling (3.3.5), we have the F-vector space of pure (trace 0) elements of B given by $B^0 = \{1\}^\perp$. The standard involution restricted to B^0 is given by $\overline{\alpha} = -\alpha$ for $\alpha \in B^0$, so equivalently B^0 is the -1-eigenspace for $\overline{}$. We have $B^0 = Fi \oplus Fj \oplus Fij$ and in this basis

$$\operatorname{nrd}|_{B^0} \simeq \langle -a, -b, ab \rangle \tag{5.2.3}$$

so that $\operatorname{disc}(\operatorname{nrd}|_{B^0}) = (ab)^2 = 1 \in F^\times/F^{\times 2}$ (cf. Example 4.3.8).

Proposition 5.2.4. *Let B, B' be quaternion algebras over F. Then the following are equivalent:*

(i) *$B \simeq B'$ are isomorphic as F-algebras;*
(ii) *$B \simeq (B')^{\mathrm{op}}$ are isomorphic as F-algebras;*
(iii) *$B \simeq B'$ are isometric as quadratic spaces; and*
(iv) *$B^0 \simeq (B')^0$ are isometric as quadratic spaces.*

If $f\colon B^0 \xrightarrow{\sim} (B')^0$ is an isometry, then f extends uniquely to either an isomorphism $f\colon B \xrightarrow{\sim} B'$ or an isomorphism $f\colon B \xrightarrow{\sim} (B')^{\mathrm{op}}$ of F-algebras.

Proof. We follow Lam [Lam2005, Theorem III.2.5]. The equivalence (i) ⇔ (ii) follows from postcomposing with the standard involution $\overline{}\colon B' \xrightarrow{\sim} (B')^{\mathrm{op}}$.

The implication (i) ⇒ (iii) follows from the fact that the standard involution on an algebra is unique and the reduced norm is determined by this standard involution, so the reduced norm on B is identified with the reduced norm on B'.

The implication (iii) ⇒ (iv) follows from Witt cancellation (Theorem 4.2.22); and (iv) ⇒ (iii) is immediate, since $B = \langle 1 \rangle \boxplus B^0$ and $B' = \langle 1 \rangle \boxplus (B')^0$ so the isometry extends by mapping $1 \mapsto 1$. (Or use Witt extension, Theorem 4.2.22(b).)

So finally we prove (iv) ⇒ (i). Let $f\colon B^0 \to (B')^0$ be an isometry of quadratic spaces. Suppose $B \simeq (a, b \mid F)$. Since f is an isometry, $\operatorname{nrd}(f(i)) = \operatorname{nrd}(i) = -a$ and

$$\operatorname{nrd}(f(i)) = f(i)\overline{f(i)} = -f(i)^2$$

so $f(i)^2 = a$. Similarly $f(j)^2 = b$. Finally, $ji = -ij$ since i, j are orthogonal (as in the proof of Main Theorem 4.4.1), but then $f(i), f(j)$ are orthogonal as well and so $f(j)f(i) = -f(i)f(j)$.

Similarly, we know that ij is orthogonal to i, j, thus $f(ij)$ is orthogonal to both $f(i)$ and $f(j)$ and so $f(ij) = uf(i)f(j)$ for some $u \in F^\times$; taking reduced norms gives $\operatorname{nrd}(ij) = u^2 \operatorname{nrd}(i)\operatorname{nrd}(j)$ so $u^2 = 1$ thus $u = \pm 1$. If $u = 1$, then $f(ij) = f(i)f(j)$, and f extends via $f(1) = 1$ to an F-algebra isomorphism $B \xrightarrow{\sim} B'$. Otherwise, $u = -1$ and $f(ij) = -f(i)f(j) = f(j)f(i)$, in which case f extends to an F-algebra *anti*-isomorphism, or equivalently an F-algebra isomorphism $B \xrightarrow{\sim} (B')^{\mathrm{op}}$; but then postcomposing with the standard involution we obtain an F-algebra isomorphism $B \xrightarrow{\sim} B'$. ∎

Main Theorem 5.2.5. *Let F be a field with* $\operatorname{char} F \neq 2$*. Then the functor $B \mapsto \operatorname{nrd}|_{B^0}$ yields an equivalence of categories between*

Quaternion algebras over F,
under F-algebra isomorphisms and anti-isomorphisms

and

Ternary quadratic forms over F with discriminant $1 \in F^\times/F^{\times 2}$,
under isometries.

Proof. The association $B \mapsto \operatorname{nrd}|_{B^0}$ gives a functor from quaternion algebras to nondegenerate ternary quadratic forms with discriminant 1, by 5.2.2; the map sends isomorphisms and anti-isomorphisms to isometries and vice versa by Proposition 5.2.4. Therefore the functor is fully faithful. To conclude, we show that the functor is essentially surjective. Let V be a nondegenerate ternary quadratic space with discriminant $1 \in F^\times/F^{\times 2}$. Choose a normalized basis for V, so that $Q \simeq \langle -a, -b, c \rangle$ with $a, b, c \in F^\times$. By hypothesis, we have $\operatorname{disc}(Q) = abc \in F^{\times 2}$, so applying the isometry rescaling the third basis vector we may suppose $c = ab$. We then associate to V the isomorphism class of the quaternion algebra $(a, b \mid F)$. The result follows. ∎

Corollary 5.2.6. *The map $B \mapsto \operatorname{nrd}|_{B^0}$ yields a bijection*

$$\left\{ \begin{array}{c} \text{Quaternion algebras over } F \\ \text{up to isomorphism} \end{array} \right\} \leftrightarrow \left\{ \begin{array}{c} \text{Ternary quadratic forms over } F \\ \text{with discriminant } 1 \in F^\times/F^{\times 2} \\ \text{up to isometry} \end{array} \right\}$$
$$\leftrightarrow \left\{ \begin{array}{c} \text{Nondegenerate ternary} \\ \text{quadratic forms over } F \\ \text{up to similarity} \end{array} \right\}$$

that is functorial with respect to F.

By the expression *functorial with respect to F*, we mean that this bijection respects (is compatible with) field extensions: explicitly, if $F \hookrightarrow K$ is an inclusion of fields, and B is a quaternion algebra with associated ternary quadratic form $Q\colon B^0 \to F$, then the quaternion algebra $B_K = B \otimes_F K$ has associated ternary quadratic form $Q_K\colon B^0_K = B^0 \otimes_F K \to K$.

Proof of Corollary 5.2.6. Functoriality boils down to the fact that

$$(B_K)^0 = (B \otimes_F K)^0 = B^0 \otimes_F K$$

for $F \hookrightarrow K$ an inclusion of fields. The first bijection is an immediate consequence of Main Theorem 5.2.5. We do not need anti-isomorphisms once we restrict to classes, since if there is an anti-isomorphism $B \xrightarrow{\sim} B'$ then composing with the standard involution gives a straight up isomorphism.

Next, we examine the natural map from isometry classes to similarity classes and show it is surjective. Every nondegenerate ternary quadratic form (or any quadratic form in odd dimension) is similar to a unique isometry class of quadratic forms with trivial discriminant: if $Q = \langle a, b, c\rangle$ with $a, b, c \in F^\times$, then $\operatorname{disc}(\langle a, b, c\rangle) = abc$ and

$$Q = \langle a, b, c\rangle \sim abc\langle a, b, c\rangle = \langle a^2bc, ab^2c, abc^2\rangle \simeq \langle bc, ac, ab\rangle$$

and $\operatorname{disc}(\langle bc, ac, ab\rangle) = (abc)^2 = 1 \in F^\times/F^{\times 2}$. Therefore the map is surjective.

To conclude, we show this map is injective. Suppose that Q, Q' are forms of discriminant 1, so $\det T, \det T' \in F^{\times 2}$. Suppose there is a similarity $Q \sim Q'$, so $Q'(f(x)) = uQ(x)$ for some $f\colon V \to V'$ and $u \in F^\times$; we show in fact that $Q \simeq Q'$ are isometric. By Exercise 4.10, we have $\det T' = u^3 \det T$, and $u = c^2 \in F^{\times 2}$. Therefore

$$Q'(c^{-1}f(x)) = c^{-2}Q'(f(x)) = u^{-1}Q'(f(x)) = Q(x)$$

and $c^{-1}f\colon V \xrightarrow{\sim} V'$ is the sought after isometry. □

Remark 5.2.7. We will refine Main Theorem 5.2.5 in section 5.6 by restricting the isometries to those that preserve orientation.

5.3 Clifford algebras

In this section, we define a functorial inverse to $B \mapsto \mathrm{nrd}|_{B^0} = Q$ in Main Theorem 5.2.5: this is the even Clifford algebra of Q. The Clifford algebra is useful in many contexts, so we define it more generally. Loosely speaking, the Clifford algebra of a quadratic form Q is the algebra generated by V subject to the condition $x^2 = Q(x)$ for all $x \in V$, so the multiplication on the Clifford algebra is induced by the quadratic form.

Let $Q\colon V \to F$ be a quadratic form with $\dim_F V = n < \infty$; in this section, we pause our assumption and allow F of arbitrary characteristic.

Proposition 5.3.1. *There exists an F-algebra* $\mathrm{Clf}\, Q$ *with the following properties:*

(i) *There is an F-linear map $\iota\colon V \to \mathrm{Clf}\, Q$ such that $\iota(x)^2 = Q(x)$ for all $x \in V$; and*

(ii) $\mathrm{Clf}\, Q$ *has the following universal property: if A is an F-algebra and $\iota_A : V \to A$ is a map such that $\iota_A(x)^2 = Q(x)$ for all $x \in V$, then there exists a unique F-algebra homomorphism $\phi : \mathrm{Clf}\, Q \to A$ such that the diagram*

$$\begin{array}{ccc} V & \xrightarrow{\ \iota\ } & \mathrm{Clf}\, Q \\ & {\scriptstyle \iota_A} \searrow & \downarrow {\scriptstyle \phi} \\ & & A \end{array}$$

commutes.

The pair $(\mathrm{Clf}\, Q, \iota)$ *is unique up to unique isomorphism.*

The algebra $\mathrm{Clf}\, Q$ in Proposition 5.3.1 is called the **Clifford algebra** of Q.

Proof. Let

$$\mathrm{Ten}\, V := \bigoplus_{d=0}^{\infty} V^{\otimes d} \tag{5.3.2}$$

where

$$V^{\otimes d} := \underbrace{V \otimes \cdots \otimes V}_{d} \quad \text{and} \quad V^{\otimes 0} := F,$$

so that

$$\mathrm{Ten}\, V = F \oplus V \oplus (V \otimes V) \oplus \dots.$$

Then $\mathrm{Ten}\, V$ has a multiplication given by tensor product: for $x \in V^{\otimes d}$ and $y \in V^{\otimes e}$ we define

$$x \cdot y = x \otimes y \in V^{\otimes(d+e)}$$

(concatenate, and possibly distribute, tensors). In this manner, $\mathrm{Ten}\, V$ has the structure of an F-algebra, and we call $\mathrm{Ten}\, V$ the **tensor algebra** of V.

Let

$$I(Q) = \langle x \otimes x - Q(x) : x \in V \rangle \subseteq \mathrm{Ten}\, V \tag{5.3.3}$$

be the two-sided ideal generated the elements $x \otimes x - Q(x)$ for all $x \in V$. Let

$$\operatorname{Clf} Q = \operatorname{Ten} V / I(Q). \tag{5.3.4}$$

The algebra $\operatorname{Clf} Q$ by construction satisfies (i). And if $\iota_A : V \to A$ is as in (ii), then the map $\iota(x) \mapsto \iota_A(x)$ for $x \in V$ extends to a unique F-algebra map $\operatorname{Ten} V \to A$; since further $\iota_A(x)^2 = Q(x)$ for all $x \in V$, this algebra map factors through $\phi\colon \operatorname{Clf} Q \to A$. By abstract nonsense (taking $A = \operatorname{Clf} Q$), we see that any other algebra having the same property as $\operatorname{Clf}(Q)$ is uniquely isomorphic to it, i.e., $\operatorname{Clf} Q$ is unique up to unique isomorphism. □

Example 5.3.5. If $Q\colon F \to F$ is the quadratic form $Q(x) = ax^2$ with $a \in F$, then $\operatorname{Clf}(F) \simeq F[x]/(x^2 - a)$ (Exercise 5.6).

Example 5.3.6. In the extreme case where $Q = 0$ identically, $\operatorname{Clf} Q \simeq \bigoplus_{d=0}^{n} \bigwedge^d V$ is canonically identified with the exterior algebra on V.

5.3.7. Let $x, y \in V$. Then in $\operatorname{Clf} Q$, we have

$$\begin{aligned} (x+y)\otimes(x+y) - x\otimes x - y\otimes y &= Q(x+y) - Q(x) - Q(y) \\ x\otimes y + y\otimes x &= T(x,y). \end{aligned} \tag{5.3.8}$$

In particular, x, y are orthogonal if and only if $x \otimes y = -y \otimes x$.

5.3.9. Let $e_1, \dots, e_n$ be an F-basis for V. Then finite tensors on these elements are an F-basis for $\operatorname{Ten} V$. In $\operatorname{Clf} Q$, by 5.3.7 we have $e_i \otimes e_i = Q(e_i)$ and $e_j \otimes e_i = T(e_i, e_j) - e_i \otimes e_j$, so an F-spanning set for $\operatorname{Clf} Q$ is given by $e_{i_1} \otimes \dots \otimes e_{i_d}$ with $1 \le i_1 < i_2 < \dots < i_d \le n$ (including 1 arising from the empty tensor product), and so

$$\dim_F \operatorname{Clf}(Q) \le \sum_{d=0}^{n} \binom{n}{d} = 2^n. \tag{5.3.10}$$

It is customary to abbreviate $e_{i_1} \otimes \dots \otimes e_{i_d} = e_{i_1} \cdots e_{i_d}$.

Example 5.3.11. If $Q \simeq \langle a_1, \dots, a_n \rangle$ is diagonal in the basis e_i, then

$$(e_{i_1} \cdots e_{i_d})^2 = \operatorname{sgn}(i_1 \ \dots \ i_d) e_{i_1}^2 (e_{i_2} \cdots e_{i_d})^2 = \dots = (-1)^{d(d-1)/2} a_{i_1} \cdots a_{i_d}.$$

Example 5.3.12. Suppose $\operatorname{char} F \neq 2$ and let $Q\colon F^2 \to F$ be the quadratic form $Q(x) = \langle a, b \rangle$. Then by a direct calculation using 5.3.9, we find

$$\operatorname{Clf} Q = F \oplus Fe_1 \oplus Fe_2 \oplus Fe_1e_2 \tag{5.3.13}$$

with multiplication $e_1^2 = a$ and $e_2^2 = b$ and $e_2e_1 = -e_1e_2$, i.e., with $i := e_1$ and $j := e_2$ we have identified $\operatorname{Clf} Q \simeq \left(\dfrac{a, b}{F}\right)$ when $a, b \neq 0$.

Example 5.3.12 generalizes as follows.

Lemma 5.3.14. *The map* $\iota\colon V \to \operatorname{Clf} Q$ *is injective, and* $\dim_F \operatorname{Clf}(Q) = 2^n$.

Proof. By Lemma 4.3.1, we may choose a basis $e_1, \dots, e_n$ for V in which $Q \simeq \langle a_1, \cdots, a_n \rangle$ is diagonal. Let A be the F-vector space with basis the symbols $z_{i_1} \cdots z_{i_d}$ for $1 \le i_1 < i_2 < \cdots < i_d \le n$. In the same way as what was considered for the relations (2.2.2), we verify directly that there is a unique, associative multiplication on A such that $z_i^2 = a_i$ and $z_j z_i = -z_i z_j$. (Alternatively, this can be viewed as a graded tensor product; see Exercise 5.21.) The map $\iota_A\colon V \to A$ by $e_i \mapsto z_i$ has $\iota_A(x)^2 = \iota_A(\sum_i x_i e_i)^2 = a_1 x_1^2 + \cdots + a_n x_n^2 = Q(x)$, so by the universal property of $\operatorname{Clf}(Q)$, there exists a unique F-algebra homomorphism $\phi\colon \operatorname{Clf}(Q) \to A$ such that $\phi\iota = \iota_A$. Since the elements $z_{i_1} \cdots z_{i_d}$ are F-linearly independent in A, so too are their preimages $e_{i_1} \cdots e_{i_d}$ in $\operatorname{Clf} Q$, so the spanning set given in 5.3.9 is in fact a basis and ϕ is an isomorphism. □

As it will cause no confusion, we may identify V with its image $\iota(V) \hookrightarrow \operatorname{Clf} Q$.

5.3.15. The reversal map, given by

$$\begin{aligned} \operatorname{rev}\colon \operatorname{Clf} Q &\to \operatorname{Clf} Q \\ x_1 \otimes \cdots \otimes x_r &\mapsto x_r \otimes \cdots \otimes x_1 \end{aligned} \tag{5.3.16}$$

on pure tensors (and extended F-linearly) is well-defined, as it maps the ideal $I(Q)$ to itself, and so it defines an involution on $\operatorname{Clf} Q$ that we call the **reversal involution**.

Lemma 5.3.17. *The association* $Q \mapsto \operatorname{Clf} Q$ *induces a faithful functor from the category of*

$$\text{quadratic forms over } F, \text{ under isometries}$$

to the category of

$$\text{finite-dimensional } F\text{-algebras with involution, under isomorphisms.}$$

Proof. Let $Q'\colon V' \to F$ be another quadratic form and let $f\colon V \to V'$ be an isometry. Then f induces an F-algebra map $\operatorname{Ten} V \to \operatorname{Ten}(V')$ and

$$f(x \otimes x - Q(x)) = f(x) \otimes f(x) - Q(x) = f(x) \otimes f(x) - Q'(f(x))$$

so f also induces an F-algebra map $\operatorname{Clf} Q \to \operatorname{Clf}(Q')$. Repeating with the inverse map, and applying the universal property, we see that these maps are inverse, so define isomorphisms. The functor is faithful because $V \subset \operatorname{Clf} Q$, so if $f\colon V \xrightarrow{\sim} V$ acts as the identity on $\operatorname{Clf} Q$ then it acts as the identity on V, so f itself is the identity. (This can be rephrased in terms of the universal property: see Exercise 5.13.) □

5.3.18. The tensor algebra $\operatorname{Ten} V$ has a natural $\mathbb{Z}_{\geq 0}$ grading by degree, and by construction (5.3.4), the quotient $\operatorname{Clf} Q = \operatorname{Ten} V/I(Q)$ retains a $\mathbb{Z}/2\mathbb{Z}$-grading

$$\operatorname{Clf} Q = \operatorname{Clf}^0 Q \oplus \operatorname{Clf}^1 Q$$

where $\operatorname{Clf}^0 Q \subseteq \operatorname{Clf} Q$ is the F-subalgebra of terms of even degree and $\operatorname{Clf}^1 Q$ the $\operatorname{Clf}^0 Q$-bimodule of terms with odd degree. The reversal involution 5.3.15 preserves $\operatorname{Clf}^0 Q$ and so descends to an involution on $\operatorname{Clf}^0 Q$.

We call $\operatorname{Clf}^0 Q$ the **even Clifford algebra** and $\operatorname{Clf}^1 Q$ the **odd Clifford bimodule** of Q. The former admits the following direct construction: let

$$\operatorname{Ten}^0 V := \bigoplus_{d=0}^{\infty} V^{\otimes 2d}$$

and let $I^0(Q) := I(Q) \cap \operatorname{Ten}^0 V$; then $\operatorname{Clf}^0 Q \simeq \operatorname{Ten}^0 V/I^0(Q)$.

5.3.19. Referring to 5.3.9, the elements $e_1e_2, \dots, e_{n-1}e_n$ generate $\operatorname{Clf}^0 Q$ as an F-algebra, and $\operatorname{Clf}^0 Q$ has basis $e_{i_1} \cdots e_{i_d}$ where d is even (including the empty product 1), so $\dim_F \operatorname{Clf}^0(Q) = 2^{n-1}$.

Example 5.3.20. Continuing Example 5.3.12, we see that the reversal involution fixes i, j and acts as the standard involution on $\operatorname{Clf}^0 Q$. So the algebra $\operatorname{Clf} Q$ is not just a quaternion algebra, but one retaining a $\mathbb{Z}/2\mathbb{Z}$-grading.

Lemma 5.3.21. *The association $Q \mapsto \operatorname{Clf}^0 Q$ defines a functor from the category of*

quadratic forms over F, under similarities

to the category of

finite-dimensional F-algebras with involution, under isomorphisms.

Proof. Let $Q' : V' \to F$ be another quadratic form and let (f, u) be a similarity, with $f : V \to V'$ and $u \in F^\times$, so that $uQ(x) = Q'(f(x))$ for all $x \in V$. We modify the proof in Lemma 5.3.17: we define a map

$$\begin{aligned} \operatorname{Ten}^0 V &\to \operatorname{Ten}^0(V') \\ x_1 \otimes \cdots \otimes x_d &\mapsto (u^{-1})^{d/2} f(x_1) \otimes \cdots \otimes f(x_d). \end{aligned}$$

Then under this map, we have

$$x \otimes x - Q(x) \mapsto u^{-1}(f(x) \otimes f(x)) - Q(x) = u^{-1}(f(x) \otimes f(x) - Q'(f(x)))$$

so $I^0(Q)$ maps to $I^0(Q')$, and the induced map $\operatorname{Clf}^0 Q \to \operatorname{Clf}^0(Q')$ is an F-algebra isomorphism. ◻

5.3.22. Note that unlike the Clifford functor, the even Clifford functor need not be faithful: for example, the map $-1 : F^2 \to F^2$ has $e_1e_2 \mapsto (-e_1)(-e_2) = e_1e_2$ so acts by the identity on $\operatorname{Clf}^0 Q$.

We now come to the important immediate application.

5.3.23. Suppose that $\operatorname{char} F \neq 2$ and let $Q(x) = \langle a, b, c\rangle$ be a nondegenerate ternary quadratic form. Then the even Clifford algebra $\operatorname{Clf}^0 Q$ is given by

$$\operatorname{Clf}^0 Q = F \oplus Fi \oplus Fj \oplus Fij$$

where $i = e_1e_2$, $j = e_2e_3$, subject to the multiplication

$$i^2 = -ab, \quad j^2 = -bc, \quad ij + ji = 0.$$

So

$$\operatorname{Clf}^0 Q \simeq \left(\frac{-ab, -bc}{F}\right).$$

Letting $k = e_3e_1$, we obtain symmetrically with the other two pairs of generators j, k or k, i that

$$\operatorname{Clf}^0 Q \simeq \left(\frac{-bc, -ac}{F}\right) \simeq \left(\frac{-ac, -ab}{F}\right).$$

The reversal involution is the standard involution on $\operatorname{Clf}^0 Q$. Letting $B = \operatorname{Clf}^0 Q$,

$$\operatorname{nrd}|_{B^0} = \langle ab, bc, ac\rangle \simeq \langle abc^2, a^2bc, ab^2c\rangle = abc\langle a, b, c\rangle.$$

So if $\operatorname{disc} Q(x) = abc \in F^{\times 2}$, then $\operatorname{nrd}|_{B^0}$ is isometric to Q. In a similar way, if $B = \left(\frac{a, b}{F}\right)$, then in Main Theorem 5.2.5 we associate $Q = \operatorname{nrd}_{B^0} = \langle -a, -b, ab\rangle$, and

$$\operatorname{Clf}^0 Q \simeq \left(\frac{-ab, ab^2}{F}\right) \simeq \left(\frac{a, b}{F}\right). \tag{5.3.24}$$

This gives another tidy proof of the bijection in Corollary 5.2.6.

Remark 5.3.25. The even Clifford map does not furnish an equivalence of categories for the same reason as in 5.3.22; one way to deal with this issue is to restrict the isometries to those that preserve orientation: we carry this out in section 5.6.

5.4 Splitting

The moral of Main Theorem 5.2.5 is that the problem of classifying quaternion algebras depends on the theory of ternary quadratic forms over that field (and vice versa). We now pursue the first consequence of this moral, and we characterize the matrix ring among quaternion algebras. Suppose that $\operatorname{char} F \neq 2$, but still $Q\colon V \to F$ a quadratic form with $\dim_F V < \infty$.

Definition 5.4.1. The **hyperbolic plane** is the quadratic form $H\colon F^2 \to F$ defined by $H(x, y) = xy$. A quadratic form is a **hyperbolic plane** if it is isometric to H.

A hyperbolic plane H is universal, its associated bilinear form has Gram matrix $\begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$ in the standard basis, and H has normalized form $H \simeq \langle 1, -1\rangle$.

Lemma 5.4.2. *Suppose Q is nondegenerate. Then Q is isotropic if and only if there exists an isometry $Q \simeq H \boxplus Q'$ with Q' nondegenerate and H a hyperbolic plane.*

Proof. For the implication ($\Leftarrow$), we have an isotropic vector from either one of the two basis vectors. For the implication ($\Rightarrow$), let $x \in V$ be isotropic, so $x \neq 0$ and satisfy $Q(x) = 0$. Since Q is nondegenerate, there exists $y \in V$ such that $T(x, y) \neq 0$; rescaling y, we may assume $T(x, y) = 1$. If $T(y, y) = 2Q(y, y) \neq 0$, replacing $y \leftarrow y - 2x/T(y, y)$ gives $T(y, y) = 0$. Thus Q restricted to $Fx + Fy$ is isometric to H, and in particular is nondegenerate. Therefore letting $V' := (Fx + Fy)^\perp$ and $Q' := Q|_{V'}$, we have $V \simeq (Fx + Fy) \boxplus V'$ and $Q \simeq H \boxplus Q'$. □

Lemma 5.4.3. *Suppose Q is nondegenerate and let $a \in F^\times$. Then the following are equivalent:*

(i) *Q represents a;*
(ii) *$Q \simeq \langle a \rangle \boxplus Q'$ for some nondegenerate form Q'; and*
(iii) *$\langle -a \rangle \boxplus Q$ is isotropic.*

Proof. For (i) $\Rightarrow$ (ii), we take $Q' = Q|_W$ and $W = \{v\}^\perp \subset V$ where $Q(v) = a$. For (ii) $\Rightarrow$ (iii), we note that $\langle -a \rangle \boxplus Q \simeq \langle a, -a \rangle \boxplus Q'$ is isotropic. For (iii) $\Rightarrow$ (i), suppose $(\langle -a \rangle \boxplus Q)(v) = 0$, so $Q(v) = ax^2$ for some $x \in F$. If $x = 0$, then Q is isotropic and by Lemma 5.4.2 represents a; if $x \neq 0$, then by homogeneity $Q(v/x) = a$ and again Q represents a. □

We now come to a main result.

Main Theorem 5.4.4. *Let $B = \left(\dfrac{a, b}{F}\right)$ be a quaternion algebra over F (with $\operatorname{char} F \neq 2$). Then the following are equivalent:*

(i) *$B \simeq \left(\dfrac{1, 1}{F}\right) \simeq \mathrm{M}_2(F)$;*
(ii) *B is not a division ring;*
(iii) *The quadratic form $\operatorname{nrd} \simeq \langle 1, -a, -b, ab \rangle$ is isotropic;*
(iv) *The quadratic form $\operatorname{nrd}|_{B^0} \simeq \langle -a, -b, ab \rangle$ is isotropic;*
(v) *The binary form $\langle a, b \rangle$ represents 1;*
(vi) *$b \in \mathrm{Nm}_{K|F}(K^\times)$ where $K = F[i]$; and*
(vi′) *$b \in \mathrm{Nm}_{K|F}(K^\times)$ where $K = F(\sqrt{a})$.*

Condition (vi) holds if and only if there exist $x, y \in F$ such that $x^2 - ay^2 = b$; if K is not a field then $K \simeq F \times F$ and $\mathrm{Nm}_{K|F}(K^\times) = F^\times$. In condition (vi′), we take $F(\sqrt{a})$ to be a splitting field for $x^2 - a$ over F, so equal to F if $a \in F^\times$. (Depending on the circumstances, one of these formulations may be more natural than the other.)

Proof. We follow Lam [Lam2005, Theorem 2.7]. The isomorphism $(1, 1 \mid F) \simeq \mathrm{M}_2(F)$ in (i) follows from Example 2.2.4. The implication (i) $\Rightarrow$ (ii) is clear. The equivalence (ii) $\Leftrightarrow$ (iii) follows from the fact that $\alpha \in B^\times$ if and only if $\operatorname{nrd}(\alpha) \in F^\times$ (Exercise 3.5).

We now prove (iii) $\Rightarrow$ (iv). Let $0 \neq \alpha \in B$ be such that $\operatorname{nrd}(\alpha) = 0$. If $\operatorname{trd}(\alpha) = 0$, then we are done. Otherwise, $\operatorname{trd}(\alpha) \neq 0$. Let β be orthogonal to $1, \alpha$, so that $\operatorname{trd}(\alpha\beta) = 0$. We cannot have both $\alpha\beta = 0$ and $\overline{\alpha}\beta = (\operatorname{trd}(\alpha) - \alpha)\beta = 0$, so we may suppose $\alpha\beta \neq 0$. But then $\operatorname{nrd}(\alpha\beta) = \operatorname{nrd}(\alpha)\operatorname{nrd}(\beta) = 0$ as desired.

To complete the equivalence of the first four we prove (iv) $\Rightarrow$ (i). Let $\beta \in B^0$ satisfy $\operatorname{nrd}(\beta) = 0$. Since $\operatorname{nrd}|_{B^0}$ is nondegenerate, there exists $0 \neq \alpha \in B^0$ such that $\operatorname{trd}(\alpha\overline{\beta}) \neq 0$. Therefore, the restriction of nrd to $F\alpha \oplus F\beta$ is nondegenerate and isotropic. By Lemma 5.4.2, we conclude there exists a basis for B^0 such that $\operatorname{nrd}|_{B^0} \simeq \langle 1, -1\rangle \boxplus \langle c\rangle = \langle 1, -1, c\rangle$; but $\operatorname{disc}(\operatorname{nrd}|_{B^0}) = -c \in F^{\times 2}$ by 5.2.2; rescaling, we may suppose $c = -1$. But then by Proposition 5.2.4 we have $B \simeq (1, 1 \mid F)$.

Now we show (iv) $\Rightarrow$ (v). For $\alpha \in B^0$,

$$\operatorname{nrd}(\alpha) = \operatorname{nrd}(xi + yj + zij) = -ax^2 - by^2 + abz^2$$

as in 5.2.2. Suppose $\operatorname{nrd}(\alpha) = 0$. If $z = 0$, then the binary form $\langle a, b\rangle$ is isotropic so is a hyperbolic plane by Lemma 5.4.2 and thus represents 1. If $z \neq 0$ then

$$a\left(\frac{y}{az}\right)^2 + b\left(\frac{x}{bz}\right)^2 = 1.$$

Next we prove (v) $\Rightarrow$ (vi). If $a \in F^{\times 2}$ then $K \simeq F \times F$ and $\operatorname{Nm}_{K|F}(K^\times) = F^\times \ni b$. If $a \notin F^{\times 2}$, then given $ax^2 + by^2 = 1$ we must have $y \neq 0$ so

$$\left(\frac{1}{y}\right)^2 - a\left(\frac{x}{y}\right)^2 = \operatorname{Nm}_{K|F}\left(\frac{1 - x\sqrt{a}}{y}\right) = b.$$

In the equivalence (vi) $\Leftrightarrow$ (vi′), the two statements are identical if $a \notin F^{\times 2}$ and both automatically satisfied if $a \in F^{\times 2}$.

To conclude, we prove (vi) $\Rightarrow$ (iii). If $b = x^2 - ay^2 \in \operatorname{Nm}_{K|F}(K^\times)$, then $\alpha = x + yi + j \neq 0$ has $\operatorname{nrd}(\alpha) = x^2 - ay^2 - b = 0$. □

We give a name to the equivalent conditions in Main Theorem 5.4.4.

Definition 5.4.5. A quaternion algebra B over F is **split** if $B \simeq \mathrm{M}_2(F)$. A field K containing F is a **splitting field** for B if $B \otimes_F K$ is split.

Example 5.4.6. The fundamental example of a splitting field for a quaternion algebra is that $\mathbb{C}$ splits the real Hamiltonians $\mathbb{H}$: we have $\mathbb{H} \otimes_{\mathbb{R}} \mathbb{C} \simeq \mathrm{M}_2(\mathbb{C})$ as in (2.4.1).

Lemma 5.4.7. *Let $K \supset F$ be a quadratic extension of fields. Then K is a splitting field for B if and only if there is an injective F-algebra homomorphism $K \hookrightarrow B$.*

Proof. First, suppose $\iota\colon K \hookrightarrow B$. We may suppose that $K = F(\sqrt{d})$ with $d \in F^\times$. Let $\mu = \iota(\sqrt{d})$, so $\mu^2 = d$. Then $1 \otimes \sqrt{d} - \mu \otimes 1$ is a zerodivisor in $B \otimes_F K$:

$$(1 \otimes \sqrt{d} - \mu \otimes 1)(1 \otimes \sqrt{d} + \mu \otimes 1) = 1 \otimes d - d \otimes 1 = 0.$$

By Main Theorem 5.4.4, we conclude that $B \otimes_F K \simeq \mathrm{M}_2(K)$.

Next we prove the converse. If $B \simeq \mathrm{M}_2(F)$ already, then any quadratic field K embeds in B (take a matrix in rational normal form) and $B \otimes_F K \simeq \mathrm{M}_2(K)$ for any K. So by Main Theorem 5.4.4, we may suppose B is a division ring. Let $K = F(\sqrt{d})$. We have $B \otimes_F K \simeq \mathrm{M}_2(K)$ if and only if $\langle -a, -b, ab \rangle$ is isotropic over K, which is to say there exist $x, y, z, u, v, w \in F$ such that

$$-a(x + u\sqrt{d})^2 - b(y + v\sqrt{d})^2 + ab(z + w\sqrt{d})^2 = 0. \tag{5.4.8}$$

Let $\alpha = xi + yj + zij$ and $\beta = ui + vj + wij$. Then $\operatorname{trd}(\alpha) = \operatorname{trd}(\beta) = 0$. Expansion of (5.4.8) (Exercise 5.14) shows that α is orthogonal to β, so $\operatorname{trd}(\alpha\beta) = 0$, and that $\operatorname{nrd}(\alpha) + d \operatorname{nrd}(\beta) = 0$. Since B is a division ring, if $\operatorname{nrd}(\beta) = c = 0$ then $\beta = 0$ so $\operatorname{nrd}(\alpha) = 0$ as well and $\alpha = 0$, a contradiction. So $\operatorname{nrd}(\beta) \neq 0$, and the element $\gamma = \alpha\beta^{-1} = c^{-1}\alpha\beta \in B$ has $\operatorname{nrd}(\gamma) = -d$ and $\operatorname{trd}(\gamma) = c^{-1}\operatorname{trd}(\alpha\beta) = 0$ so $\gamma^2 = d$ as desired. □

Example 5.4.9. If $B = \left(\dfrac{a, b}{F}\right)$, then either $a \in F^{\times 2}$ and $B \simeq \left(\dfrac{1, b}{F}\right) \simeq \mathrm{M}_2(F)$ is split, or $a \notin F^{\times 2}$ and $K = F(\sqrt{a})$ splits B.

Example 5.4.10. Let p be an odd prime and let a be a quadratic nonresidue modulo p. We claim that $\left(\dfrac{a, p}{\mathbb{Q}}\right)$ is a division quaternion algebra over $\mathbb{Q}$. By Main Theorem 5.4.4, it suffices to show that the quadratic form $\langle 1, -a, -p, ap \rangle$ is anisotropic. So suppose that $t^2 - ax^2 = p(y^2 - az^2)$ with $t, x, y, z \in \mathbb{Q}$ not all zero. The equation is homogeneous, so we can multiply through by a common denominator and suppose that $t, x, y, z \in \mathbb{Z}$ with $\gcd(t, x, y, z) = 1$. Reducing modulo p we find $t^2 \equiv ax^2 \pmod{p}$; since a is a quadratic nonresidue, we must have $t \equiv x \equiv 0 \pmod{p}$. Plugging back in and cancelling a factor of p we find $y^2 \equiv az^2 \equiv 0 \pmod{p}$, and again $y \equiv z \equiv 0 \pmod{p}$, a contradiction.

5.5 Conics, embeddings

Following Main Theorem 5.2.5, we are led to consider the zero locus of the quadratic form $\operatorname{nrd}|_{B^0}$ up to scaling; this gives a geometric way to view the preceding results.

Definition 5.5.1. A **conic** $C \subset \mathbb{P}^2$ over F is a nonsingular projective plane curve of degree 2. An **isomorphism** of conics C, C' over F is an element $f \in \mathrm{PGL}_3(F) = \operatorname{Aut}(\mathbb{P}^2)(F)$ that induces an isomorphism of curves $f \colon C \xrightarrow{\sim} C'$.

If we identify

$$\mathbb{P}(B^0) := (B^0 \smallsetminus \{0\})/F^\times \simeq \mathbb{P}^2(F)$$

with (the points of) the projective plane over F, then the vanishing locus $C = V(Q)$ of $Q = \operatorname{nrd}|_{B^0}$ defines a conic over F: if we take the basis i, j, ij for B^0, then the conic C is defined by the vanishing of the equation

$$Q(x, y, z) = \operatorname{nrd}(xi + yj + zij) = -ax^2 - by^2 + abz^2 = 0.$$

Here, nondegeneracy of the quadratic form is equivalent to the nonsingularity of the associated plane curve (Exercise 4.13).

The following corollary is then simply a rephrasing of Main Theorem 5.2.5.

Corollary 5.5.2. *The map* $B \mapsto C = V(\operatorname{nrd}|_{B^0})$ *yields a bijection*

$$\left\{\begin{array}{c}\text{Quaternion algebras over } F\\ \text{up to isomorphism}\end{array}\right\} \leftrightarrow \left\{\begin{array}{c}\text{Conics over } F\\ \text{up to isomorphism}\end{array}\right\}$$

that is functorial with respect to F.

Main Theorem 5.4.4 also extends to this context.

Theorem 5.5.3. *The following are equivalent:*

(i) $B \simeq \mathrm{M}_2(F)$;

(vii) *The conic* C *associated to* B *has an* F*-rational point.*

By Lemma 5.4.7, a quadratic field K over F embeds in B if and only if the ternary quadratic form $\operatorname{nrd}|_{B^0}$ represents 0 over K. We can also rephrase this in terms of the values represented by $\operatorname{nrd}|_{B^0}$.

Lemma 5.5.4. *Let* K *be a quadratic extension of* F *of discriminant* d. *Then* $K \hookrightarrow B$ *if and only if* $\operatorname{nrd}|_{B^0}$ *represents* $-d$ *over* F.

Proof. Write $K = F(\sqrt{d})$. Then $K \hookrightarrow B$ if and only if there exists $\alpha \in B$ such that $\alpha^2 = d$ if and only if there exists $\alpha \in B$ with $\operatorname{trd}(\alpha) = 0$ and $\operatorname{nrd}(\alpha) = -d$, as claimed. □

Remark 5.5.5. Two conics over F are isomorphic (as plane curves) if and only if their function fields are isomorphic (Exercise 5.23).

5.6 Orientations

To conclude, we show that the notion of orientation underlying the definition of special isometries (as in Example 4.5.5) extends more generally to isometries between two different quadratic spaces by keeping track of one bit of extra information, refining Main Theorem 5.2.5. We follow Knus–Murkurjev–Rost–Tignol [KMRT98, Theorem 15.2]. We retain our hypothesis that $\operatorname{char} F \neq 2$.

Let $Q\colon V \to F$ be a quadratic space with $\dim_F V = n$ odd.

Lemma 5.6.1. *Suppose* Q *has* signed *discriminant* $\operatorname{sgndisc} Q = d \in F^\times/F^{\times 2}$. *Let* $A := \operatorname{Clf} Q$ *be the Clifford algebra of* Q, *and let* $K = Z(A)$ *be the center of* A. *Then* $K \simeq F[x]/(x^2 - d)$.

The signed discriminant gives a simpler statement; one could equally well work with the usual discriminant and keep track of the sign.

Proof. We do the case $n = 3$. We may suppose $V \simeq F^3$ with standard basis e_1, e_2, e_3 and that $Q \simeq \langle a, b, c\rangle$ is diagonal, with $\operatorname{sgndisc}(Q) = -abc = d$. We have the relation $e_i e_j = -e_j e_i$ for $i \neq j$; for all $i = 1, 2, 3$, conjugation by e_i acts by -1 on e_j and $e_i e_j$

for $j \neq i$. This implies $Z(A) \subseteq F + Fe_1e_2e_3$. Let $\delta := e_1e_2e_3 = e_2e_3e_1 = e_3e_1e_2$; then $\delta e_i = e_i\delta$ for $i = 1, 2, 3$, so $Z(A) = F[\delta]$. We compute

$$\delta^2 = (e_1e_2e_3)(e_1e_2e_3) = e_1^2(e_2e_3)(e_2e_3) = -abc = \operatorname{sgndisc}(Q) = d. \qquad (5.6.2)$$

Therefore $K \simeq F[x]/(x^2 - d)$.

The general case is requested in Exercise 5.19: with a basis $e_1, \dots, e_n$ for V, the center is generated over F by $\delta = e_1 \cdots e_n$. ◻

From now on, suppose $\operatorname{sgndisc}(Q) = d = 1$.

Definition 5.6.3. An **orientation** of Q is a choice of $\zeta \in Z(\operatorname{Clf} Q) \smallsetminus F$ with $\zeta^2 = 1$.

5.6.4. Q has exactly two choices of orientation ζ, differing by sign, by Lemma 5.6.1: under an isomorphism $Z(\operatorname{Clf} Q) \simeq F \times F$, the two orientations are $(-1, 1)$ and $(1, -1)$. More intrinsically, given an orientation ζ, we have a projection $K \to K/(\zeta - 1) \simeq F$, and conversely given a projection $\pi : K \to F$, there is a unique orientation ζ with $\pi(\zeta) = 1$ (the other maps to -1, by F-linearity).

Definition 5.6.5. Let ζ, ζ' be orientations on Q, Q'. An isometry $f : V \to V'$ is **oriented** (with respect to ζ, ζ') if in the induced map $f : Z(\operatorname{Clf} Q) \to Z(\operatorname{Clf} Q')$ we have $f(\zeta) = \zeta'$.

5.6.6. An oriented isometry is the same as a special isometry (Definition 4.5.1) when $V \simeq F^n$ (n still odd), as follows. Let $A = \operatorname{Clf} Q$. Let $e_1, \dots, e_n$ be a basis for V adapted as in the proof of Lemma 5.6.1 and $\delta = e_1 \dots e_n$. Then $Z(A)$ is generated by δ and $\delta^2 = 1$. If $f \in \mathrm{O}(Q)(F)$, then $f(\delta) = (\det f)\delta$, so $\zeta = \pm\delta$ is preserved if and only if $\det(f) = 1$, and this is independent of the choice of orientation.

So we define the **oriented** or **special orthogonal group** of a quadratic space by choosing an orientation and letting

$$\mathrm{SO}(Q)(F) := \{f \in \mathrm{O}(Q)(F) : f \text{ is oriented}\};$$

the resulting group is independent of the choice, and we recover the same group as in Definition 4.5.1.

5.6.7. Let $B = \left(\dfrac{a, b}{F}\right)$ be a quaternion algebra over F. In previous sections, we took $\operatorname{nrd}|_{B^0} : B^0 \to F$, a nondegenerate ternary quadratic space of discriminant 1. Since we are working with the signed discriminant, we take instead $-\operatorname{nrd}|_{B^0} : B^0 \to F$ with $\operatorname{sgndisc}(-\operatorname{nrd}|_{B^0}) = 1$; this map has a nice description as the squaring map, since $\alpha^2 = -\operatorname{nrd}(\alpha)$ for $\alpha \in B^0$.

We claim that B^0 has a canonical orientation. We have an inclusion $\iota : B^0 \hookrightarrow B$ with $\iota(x)^2 = -\operatorname{nrd}(x)$ for all $x \in B^0$. By the universal property of Clifford algebras, we get an F-algebra homomorphism $\phi :\ \operatorname{Clf}(B^0) \to B$. We see that ϕ is surjective so it induces an F-algebra map $\pi : Z(\operatorname{Clf}(B^0)) \to Z(B) = F$ (Exercise 2.8). This defines a unique orientation $\zeta_B = \zeta$ with $\zeta - 1 \in \ker \pi$, by 5.6.4.

Explicitly, let i, j, k be the standard basis for B with $k = ij$. Then $\operatorname{nrd}(k) = ab$, and i, j, k is a basis for B^0. Let $\zeta = ijk^{-1} = -ijk/(ab) \in Z(\operatorname{Clf}(B^0))$. Then

$\delta^2 = -ab(-ab)/(ab)^2 = 1$ as in (5.6.2). Multiplying out in B, we get $\phi(\zeta) = 1 \in B$, so ζ is the same orientation as in the previous paragraph.

The following theorem then refines Main Theorem 5.2.5.

Theorem 5.6.8. *Let F be a field with* $\operatorname{char} F \neq 2$. *Then the functors*

$$\begin{aligned} (Q,\zeta) &\mapsto \operatorname{Clf}^0 Q \\ (-\operatorname{nrd}|_{B^0}, \zeta_B) &\mapsfrom B \end{aligned}$$

yield an equivalence of categories between

Oriented ternary quadratic forms over F with signed discriminant $1 \in F^\times/F^{\times 2}$, under oriented isometries.

and

Quaternion algebras over F, under F-algebra isomorphisms.

Proof. Let B be a quaternion algebra. As in 5.6.7, the inclusion $\iota\colon B^0 \hookrightarrow B$ gives an F-algebra homomorphism $\operatorname{Clf}(-\operatorname{nrd}|_{B^0}) \to B$ which restricts to a canonical F-algebra homomorphism $\operatorname{Clf}^0(-\operatorname{nrd}|_{B^0}) \to B$. In fact, in coordinates, this map is the isomorphism (5.3.24): choosing the standard basis i, j, k for $B = (a, b \mid F)$, and letting $e_1 = i, e_2 = j, e_3 = k$, we have

$$\operatorname{Clf}^0(-\operatorname{nrd}|_{B^0}) = \operatorname{Clf}^0(\langle a, b, -ab\rangle) = \left(\frac{-ab, ab^2}{F}\right)$$

with the standard generators $i_0 := e_1e_2 = ij$ and $j_0 := e_2e_3 = jk$. We define the isomorphism

$$\begin{aligned} \left(\frac{-ab, ab^2}{F}\right) &\to \left(\frac{a, b}{F}\right) \\ i_0, j_0 &\mapsto ij, jk. \end{aligned}$$

Therefore, the canonical isomorphism $\operatorname{Clf}^0(-\operatorname{nrd}|_{B^0}) \xrightarrow{\sim} B$ yields a natural isomorphism between these composed functors and the identity functor, giving an equivalence of categories.

Conversely, let (Q, ζ) be an oriented ternary quadratic space, let $B = \operatorname{Clf}^0 Q$, and consider $(-\operatorname{nrd}|_{B^0}, \zeta_B)$. We define a natural oriented isometry between these two spaces. We have a natural inclusion $V \hookrightarrow \operatorname{Clf} Q$, and we define the linear map

$$\begin{aligned} m_\zeta\colon V &\to B \\ v &\mapsto v\zeta; \end{aligned}$$

since $v, \zeta \in \operatorname{Clf}^1 Q$, we have $v\zeta \in \operatorname{Clf}^0 Q = B$. We now show that m_ζ induces an oriented isometry $m_\zeta : V \to B^0$. To do so, we let $V \simeq F^3$ by choosing an orthogonal basis e_1, e_2, e_3 in which $Q \simeq \langle a, b, c\rangle$ and $-abc = 1$. We identify $B \simeq (-ab, -bc \mid F)$ as

in 5.3.23, with $i = e_1e_2$ and $j = e_2e_3$, and we let $k = ij = -be_3e_1$ so $k^2 = b^2(-ac) = b$. Then $\zeta = \epsilon e_1e_2e_3$ with $\epsilon = \pm 1$, and

$$\begin{aligned} \epsilon m_\zeta(e_1) &= e_1(e_1e_2e_3) = ae_2e_3 = aj \\ \epsilon m_\zeta(e_2) &= e_2(e_1e_2e_3) = -be_1e_3 = -k \\ \epsilon m_\zeta(e_3) &= e_3(e_1e_2e_3) = ci; \end{aligned} \tag{5.6.9}$$

so in particular $m_\zeta(V) \subseteq B^0$. The map is an isometry, because

$$-\operatorname{nrd}(m_\zeta(v)) = -\operatorname{nrd}(v\zeta) = (v\zeta)^2 = v^2 = -\operatorname{nrd}(v) \tag{5.6.10}$$

since $\zeta^2 = 1$ and ζ is central. Finally, the map is oriented:

$$\begin{aligned} m_\zeta(\zeta) &= m_\zeta(\epsilon e_1e_2e_3) = \epsilon(e_1\zeta)(e_2\zeta)(e_3\zeta) \\ &= \epsilon(\epsilon aj)(-\epsilon k)(\epsilon ci) = (-ac)(ijk) = (-abc)ijk^{-1} = ijk^{-1} = \zeta_B. \end{aligned}$$

This natural oriented isometry gives a natural transformation between these composed functors and the identity functor, and the statement follows. □

Remark 5.6.11. Theorem 5.6.8 can be seen as a manifestation of the isomorphism of Dynkin diagrams $A_1 \simeq B_1$ (consisting of a single node •), corresponding to the isomorphism of Lie algebras $\mathfrak{sl}_2 \simeq \mathfrak{so}_3$. This is just one of the (finitely many) exceptional isomorphisms—the others are just as beautiful, with deep implications, and the reader is encouraged to read the bible by Knus–Merkurjev–Rost–Tignol [KMRT98, §15].

We record the following important consequence.

Corollary 5.6.12. *We have* $\operatorname{Aut}(B) \simeq B^\times/F^\times$.

Proof. We take stabilizers of objects on both sides of the equivalence of categories in Theorem 5.6.8; we find $\operatorname{Aut}(B) \simeq \operatorname{SO}(Q)(F)$ if B corresponds to Q. But by Proposition 4.5.10, there is an isomorphism $B^\times/F^\times \simeq \operatorname{SO}(Q)(F)$, and the result follows. □

Remark 5.6.13. We will return to Corollary 5.6.12 in the Skolem–Noether theorem in section 7.7, generalizing to the context of embeddings into a simple algebra.

To conclude, we extend the notion of oriented isometry to similarities.

5.6.14. Let ζ, ζ' be orientations on quadratic spaces V, V' and suppose $\dim V = \dim V' = n = 2m$ is even. Then a similarity (f, u) from V to V' induces an F-linear map $u^{-m}\bigwedge^n f \colon \bigwedge^n(V) \to \bigwedge^n(V')$, and we say (f, u) is **oriented** if the map $u^{-m}\bigwedge^n f$ preserves orientations. We define

$$\operatorname{GSO}(Q)(F) := \{(f, u) \in \operatorname{GO}(Q)(F) : (f, u) \text{ is oriented}\}$$

and recover the same group as in 4.5.4. If n is odd, we declare that every similarity is oriented and let $\operatorname{GSO}(Q)(F) := \operatorname{GO}(Q)(F)$.

Exercises

Throughout, let F be a field with char $F \neq 2$.

1. Let B, B' be quaternion algebras over F. Show that if the quadratic forms nrd_B and $\mathrm{nrd}_{B'}$ are similar, then they are isometric.
2. Consider the *hyperbolic quaternions* H_{Mac} of Macfarlane (1.2.1).

   (a) Show that H_{Mac} is the Clifford algebra of $\langle 1,1,1\rangle$ over $\mathbb{R}$.
   (a) Show that H_{Mac} is isomorphic as an algebra over $\mathbb{C} = \mathbb{R}(\sqrt{-1})$ to the even Clifford algebra of the ternary quadratic form $-\sqrt{-1}\langle 1,1,1\rangle$.

3. Prove the implication (vi) $\Rightarrow$ (v) of Main Theorem 5.4.4 directly.
4. Use Main Theorem 5.4.4(vi) to give another proof that there is no division quaternion algebra B over a finite field $F = \mathbb{F}_q$ (with q odd).
5. (a) Show that the quadratic form $Q(x,y,z) = x^2 + y^2 + z^2$ is isotropic over $\mathbb{F}_p$ for all odd primes p. Conclude that $(-1,-1 \mid \mathbb{F}_p) \simeq \mathrm{M}_2(\mathbb{F}_p)$. *[Hint: count squares and nonsquares.]*
   (b) More generally, show that every ternary quadratic form over a finite field $\mathbb{F}_q$ (with q odd) is isotropic. *[Hint: Reduce to the case of finding a solution to $y^2 = f(x)$ where f is a polynomial of degree 2.]* Use Main Theorem 5.4.4(iv) to give yet another proof that there is no division quaternion algebra B over $\mathbb{F}_q$.
   (c) Show that over a finite field $\mathbb{F}_q$ with q odd, there is a unique anisotropic binary quadratic form up to isometry.
6. Show that if $Q\colon F \to F$ is the quadratic form $Q(x) = ax^2$ with $a \in F$, then $\mathrm{Clf}(F) \simeq F[x]/(x^2 - a)$.
7. Show that $(-1, 26)_{\mathbb{Q}} = 1$, i.e., $\left(\dfrac{-1, 26}{\mathbb{Q}}\right) \simeq \mathrm{M}_2(\mathbb{Q})$.
8. Let p be prime. Show that $\left(\dfrac{-1, p}{\mathbb{Q}}\right) \simeq \mathrm{M}_2(\mathbb{Q})$ if and only if $p = 2$ or $p \equiv 1 \pmod 4$.
9. Show that
$$\left(\frac{-2,-3}{\mathbb{Q}}\right) \simeq \left(\frac{-1,-1}{\mathbb{Q}}\right) \text{ but that } \left(\frac{-2,-5}{\mathbb{Q}}\right) \not\simeq \left(\frac{-1,-1}{\mathbb{Q}}\right).$$
10. Let $B = (a, b \mid F)$ be a quaternion algebra over F. Give a constructive (algorithmic) proof of the implication (iv) $\Rightarrow$ (i) in Main Theorem 5.4.4, as follows.

    Let $\epsilon = xi + yj + zij \in B$ satisfy $\mathrm{nrd}(\epsilon) = -ax^2 - by^2 + abz^2 = -\epsilon^2 = 0$.

    (a) Show that there exists $k \in \{i, j, ij\}$ such that $\mathrm{trd}(\epsilon k) = s \neq 0$.
    (b) Let $t := \mathrm{trd}(k)$ and $n := \mathrm{nrd}(k)$, and let $\epsilon' := s^{-1}\epsilon$. Let
$$\begin{aligned} i' &:= \epsilon' k - (k+t)\epsilon' \\ j' &:= k + (-tk + n + 1)\epsilon'. \end{aligned}$$

Show that i', j' generate B as an F-algebra, and that $(i')^2 = (j')^2 = 1$ and $j'i' = -i'j'$. Conclude that $B \simeq \mathrm{M}_2(F)$.

(c) Show that $I := F\epsilon' + Fk\epsilon'$ is a left ideal of B with $\dim_F I = 2$, and interpret (b) as arising from the left multiplication map $B \to \mathrm{End}_F(I) \simeq \mathrm{M}_2(F)$.

11. Let B be a quaternion algebra over F. Let Q be the reduced norm on B, and for clarity write $e_0 = 1$, $e_1 = i$, $e_2 = j$, $e_3 = k$ as a basis for the domain of Q.

    (a) Let $C^0 = \mathrm{Clf}^0 Q$ be the even Clifford algebra of the reduced norm Q. Show that $Z(C^0) \simeq F \times F$. *[Hint: $Z(C^0)$ is generated by $e_0e_1e_2e_3$.]*
    (b) Show that $C^0 \simeq B \times B^{\mathrm{op}} (\simeq B \times B)$ as F-algebras.
    (c) Prove that if B' is a quaternion algebra over F then $B \simeq B'$ are isomorphic as F-algebras if and only if the reduced norms $Q \sim Q'$ are similar as quadratic spaces.

12. Let $Q\colon V \to F$ be a nondegenerate quadratic form. Show that the reversal map $^{-}\colon \mathrm{Clf}^0 Q \to \mathrm{Clf}^0 Q$ on the Clifford algebra has the property that $x\overline{x} \in F$ for all pure tensors $x = e_1e_2\cdots e_d$, but defines a standard involution on $\mathrm{Clf}\, Q$ if and only if $V = \{0\}$ and on $\mathrm{Clf}^0 Q$ if and only if $\dim_F V \leq 3$.

13. Give another proof of Lemma 5.3.17 using the universal property of the Clifford algebra.

▹ 14. Expand (5.4.8) and prove as a consequence that if $\alpha = xi + yj + zij$ and $\beta = ui + vj + wij$, then $\mathrm{trd}(\alpha\beta) = 0$ (so α is orthogonal to β) and moreover $\mathrm{nrd}(\alpha) + d\,\mathrm{nrd}(\beta) = 0$.

15. Show that the Hilbert symbol is Galois equivariant, in the following sense: for all field automorphisms $\sigma \in \mathrm{Aut}(F)$ and all $a, b \in F^\times$, we have $(\sigma(a), \sigma(b))_F = (a, b)_F$.

16. Let $a, b, b' \in F^\times$. Show that there exists an F-linear isomorphism $\phi\colon (a, b \mid F) \xrightarrow{\sim} (a, b' \mid F)$ with $\phi(i) = i'$ if and only if $b/b' \in \mathrm{Nm}_{K|F}(K^\times)$ where $K = F(\sqrt{a})$. [More generally, see Corollary 7.7.6.]

17. Let $a \in \mathbb{Q}^\times \smallsetminus \mathbb{Q}^{\times 2}$. Show that there are infinitely many distinct isomorphism classes of conics $x^2 - ay^2 = bz^2$ for $b \in \mathbb{Q}^\times$.

18. Let $K = F(a, b)$ with a, b algebraically independent, transcendental elements. Show that the *generic* quaternion algebra $\left(\dfrac{a, b}{K}\right)$ is a division algebra. *[Hint: show the associated ternary quadratic form is anisotropic.]*

▹ 19 Prove Lemma 5.6.1 for general odd n as follows.

    (a) For a subset $I = \{i_1, \dots, i_r\} \subseteq \{1, \dots, n\}$, let $e_I = e_{i_1} \cdots e_{i_r}$ with $i_1 < \cdots < i_r$. Then for subsets $I, J \subseteq \{1, \dots, n\}$, show that
    $$e_I e_J = e_J e_I (-1)^{\#I \cdot \#J - \#(I \cap J)}.$$
    (b) Show that $Z(\mathrm{Clf}\, Q) = F[\delta] \simeq F[x]/(x^2 - d)$ where $\delta = e_1e_2\dots e_n$ and $d = \mathrm{sgndisc}(Q)$. *[Hint: Argue on bases and choose $\#J = 2$ with $I \cap J = 1$.]*

20. Let $Q\colon V \to F$ be a quadratic form. Show that the even Clifford algebra $\operatorname{Clf}^0 Q$ with its map $\iota\colon V \otimes V \to \operatorname{Clf}^0 Q$ has the following universal property: if A is an F-algebra and $\iota_A : V \otimes V \to A$ is an F-linear map such that

    (i) $\iota_A(x \otimes x) = Q(x)$ for all $x \in V$, and
    (ii) $\iota_A(x \otimes y)\iota_A(y \otimes z) = Q(y)\iota_A(x \otimes z)$ for all $x, y, z \in V$,

    then there exists a unique F-algebra homomorphism $\phi : \operatorname{Clf}^0 Q \to A$ such that the diagram

$$\begin{array}{ccc} V \otimes V & \xrightarrow{\ \iota\ } & \operatorname{Clf}^0 Q \\ & {\scriptstyle \iota_A}\searrow & \downarrow{\scriptstyle \phi} \\ & & A \end{array}$$

    commutes. Conclude that the pair $(\operatorname{Clf}^0 Q, \iota)$ is unique up to unique isomorphism.

21. In this exercise, we consider graded tensor products, giving an alternate verification of Lemma 5.3.14.
    Let $A = A_0 \oplus A_1$ and $B = B_0 \oplus B_1$ be finite-dimensional F-algebras equipped with a $\mathbb{Z}/2\mathbb{Z}$-grading. We define the graded tensor product $A \widehat{\otimes} B$ to be the usual tensor product as an F-vector space but with multiplication law defined on simple tensors by

$$(a \otimes b) \cdot (a' \otimes b') = (-1)^{(\deg a')(\deg b)}(aa' \otimes bb').$$

    (a) Show that $A \widehat{\otimes} B$ is an F-algebra of dimension $(\dim_F A)(\dim_F B)$.
    (b) Let $Q_1 : V_1 \to F$ and $Q_2 : V_2 \to F$, and let $Q := Q_1 \boxplus Q_2$ be the orthogonal direct sum on the quadratic space $V := V_1 \boxplus V_2$. Show there is a canonical isomorphism of Clifford algebras

$$\operatorname{Clf}(Q) \cong \operatorname{Clf}(Q_1) \widehat{\otimes} \operatorname{Clf}(Q_2).$$

    (c) Observe that (b) gives another proof of Lemma 5.3.14.

22. For $i = 1, 2$, let $Q_i : V_i \to F$ be quadratic forms over F.

    (a) Prove that there exists a canonical R-algebra isomorphism

$$\operatorname{Clf}^0(Q_1 \boxplus Q_2) \xrightarrow{\sim} (\operatorname{Clf}^0(Q_1) \otimes \operatorname{Clf}^0(Q_2)) \oplus (\operatorname{Clf}^1(Q_1) \otimes \operatorname{Clf}^1(Q_2))$$

    where $\operatorname{Clf}^1(Q_1) \otimes \operatorname{Clf}^1(Q_2)$ has multiplication induced from the full Clifford algebras $\operatorname{Clf}(Q_1)$ and $\operatorname{Clf}(Q_2)$.
    (b) Prove that there is a $\operatorname{Clf}^0(Q_1 \boxplus Q_2)$-bimodule isomorphism

$$\operatorname{Clf}^1(Q_1 \boxplus Q_2) \xrightarrow{\sim} (\operatorname{Clf}^0(Q_1) \otimes \operatorname{Clf}^1(Q_2)) \oplus (\operatorname{Clf}^1(Q_1) \otimes \operatorname{Clf}^0(Q_2))$$

    with bimodule structure induced by multiplication in the full Clifford algebra.

23. In this exercise, we assume background in algebraic curves. Show that two conics over F are isomorphic (as projective plane curves) if and only if their function fields are isomorphic. *[Hint: conics are anticanonically embedded—the restriction of* $\mathscr{O}_{\mathbb{P}^2}(-1)$ *to the conic is a canonical sheaf—so an isomorphism of function fields induces an linear isomorphism of conics.]*

This chapter is licensed under the terms of the Creative Commons Attribution-NonCommercial 4.0 International License (http://creativecommons.org/licenses/by-nc/4.0/), which permits any noncommercial use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license and indicate if changes were made.

The images or other third party material in this chapter are included in the chapter's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the chapter's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.

Chapter 6

Characteristic 2

In this chapter, we extend the results from the previous four chapters to the neglected case where the base field has characteristic 2. Throughout this chapter, let F be a field with algebraic closure F^{al}.

6.1 Separability

To get warmed up, we give a different notation (symbol) for quaternion algebras that holds in any characteristic and which is convenient for many purposes.

Definition 6.1.1. Let A be a commutative, finite-dimensional algebra over F. We say A is **separable** if

$$A \otimes_F F^{\text{al}} \simeq F^{\text{al}} \times \cdots \times F^{\text{al}};$$

otherwise, we say A is **inseparable**.

Example 6.1.2. If $A \simeq F[x]/(f(x))$ with $f(x) \in F[x]$, then A is separable if and only if f has distinct roots in F^{al}.

6.1.3. If $\operatorname{char} F \neq 2$, and K is a quadratic F-algebra, then after completing the square, we see that the following are equivalent:

(i) K is separable;
(ii) $K \simeq F[x]/(x^2 - a)$ with $a \neq 0$;
(iii) K is reduced (K has no nonzero nilpotent elements);
(iv) K is a field or $K \simeq F \times F$.

6.1.4. If $\operatorname{char} F = 2$, then a quadratic F-algebra K is separable if and only if

$$K \simeq F[x]/(x^2 + x + a)$$

for some $a \in F$. A quadratic algebra of the form $K = F[x]/(x^2 + a)$ with $a \in F$ is inseparable.

Now we introduce the more general notation.

© The Author(s) 2021

J. Voight, *Quaternion Algebras*, Graduate Texts in Mathematics 288,
https://doi.org/10.1007/978-3-030-56694-4_6

6.1.5. Let K be a separable quadratic F-algebra, and let $b \in F^\times$. We denote by

$$\left(\frac{K,b}{F}\right) := K \oplus Kj$$

the F-algebra with basis $1, j$ as a left K-vector space and with the multiplication rules $j^2 = b$ and $j\alpha = \overline{\alpha} j$ for $\alpha \in K$, where $\overline{}$ is the standard involution on K (the nontrivial element of $\mathrm{Gal}(K \mid F)$ if K is a field). We will also write $(K, b \mid F)$ for formatting.

From 6.1.3, if $\operatorname{char} F \neq 2$ then writing $K \simeq F[x]/(x^2 - a)$ we see that

$$\left(\frac{K,b}{F}\right) \simeq \left(\frac{a,b}{F}\right)$$

is a quaternion algebra over F. The point is that we cannot complete the square in characteristic 2, so the more general notation gives a characteristic-independent way to define quaternion algebras. In using this symbol, we are breaking the symmetry between the standard generators i, j, but otherwise have not changed anything about the definition.

6.2 Quaternion algebras

Throughout the rest of this chapter, we suppose that $\operatorname{char} F = 2$. (We will occasionally remind the reader of this supposition, but it is meant to hold throughout.)

Definition 6.2.1. An algebra B over F (with $\operatorname{char} F = 2$) is a **quaternion algebra** if there exists an F-basis $1, i, j, k$ for B such that

$$i^2 + i = a, \quad j^2 = b, \quad \text{and} \quad k = ij = j(i+1) \tag{6.2.2}$$

with $a \in F$ and $b \in F^\times$.

Just as when $\operatorname{char} F \neq 2$, we find that the multiplication table for a quaternion algebra B is determined by the rules (6.2.2), e.g.

$$jk = j(ij) = (ij + j)j = bi + b = kj + b.$$

We denote by $\left[\frac{a,b}{F}\right)$ or $[a, b \mid F)$ the F-algebra with basis $1, i, j, ij$ subject to the multiplication rules (6.2.2). The algebra $\left[\frac{a,b}{F}\right)$ is not symmetric in a, b (explaining the choice of notation), but it is still functorial in the field F.

If we let $K = F[i] \simeq F[x]/(x^2 + x + a)$, then

$$\left[\frac{a,b}{F}\right) \simeq \left(\frac{K,b}{F}\right)$$

and our notation extends that of Section 6.1.

Example 6.2.3. The ring $\mathrm{M}_2(F)$ of 2×2-matrices with coefficients in F is again a quaternion algebra over F, via the isomorphism

$$\left[\frac{1,1}{F}\right) \xrightarrow{\sim} \mathrm{M}_2(F)$$
$$i, j \mapsto \begin{pmatrix} 0 & 1 \\ 1 & 1 \end{pmatrix}, \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}.$$

Lemma 6.2.4. *An F-algebra B is a quaternion algebra if and only if there exist F-algebra generators $i, j \in B$ satisfying*

$$i^2 + i = a, \; j^2 = b, \; \textit{and} \; ij = j(i+1). \tag{6.2.5}$$

Proof. Proven the same way as Lemma 2.2.5. □

6.2.6. Let $B = [a, b \mid F)$ be a quaternion algebra over F. Then B has a (unique) standard involution $\overline{} : B \to B$ given by

$$\alpha = t + xi + yj + zij \mapsto \overline{\alpha} = x + \alpha = (t + x) + xi + yj + zij$$

since

$$\begin{aligned} \alpha\overline{\alpha} &= (t + xi + yj + zij)((t + x) + xi + yj + zij) \\ &= t^2 + tx + ax^2 + by^2 + byz + abz^2 \in F. \end{aligned} \tag{6.2.7}$$

Consequently, one has a reduced trace and reduced norm on B as in Chapter 3.

We now state a version of Theorem 3.5.1 in characteristic 2; the proof is similar and is left as an exercise.

Theorem 6.2.8. *Let B be a division F-algebra with a standard involution that is not the identity. Then either B is a separable quadratic field extension of F or B is a quaternion algebra over F.*

Proof. Exercise 6.9. (This theorem is also implied by Theorem 6.4.1.) □

6.3 ∗ Quadratic forms

We now turn to the theory of quadratic forms over F with $\operatorname{char} F = 2$. The basic definitions from section 4.2 apply. For further reference, Grove [Grov2002, Chapters 12–14] treats quadratic forms in characteristic 2, and the book by Elman–Karpenko–Merkurjev [EKM2008, Chapters I–II] discusses bilinear forms and quadratic forms in all characteristics.

Let $Q: V \to F$ be a quadratic form with $\dim_F V = n < \infty$ and associated bilinear form T. Then $T(x, x) = 2Q(x) = 0$ for all $x \in V$, so one cannot recover the quadratic form from the symmetric (equivalently, alternating) bilinear form.

6.3.1. We begin with the definition of the discriminant. When n is even, we simply define $\operatorname{disc}(Q) = \det(T) \in F/F^{\times 2}$—this is equivalent to Definition 4.3.3 when $\operatorname{char} F \neq 2$, having absorbed the square power of 2.

When n is odd, the symmetric matrix T always has determinant 0 (Exercise 6.8); we need to "divide this by 2". So instead we work with a *generic* quadratic form, as follows. Consider the quadratic form

$$Q^{\mathrm{univ}}(x_1,\dots,x_n) := \sum_{1\le i\le j\le n} a_{ij}x_ix_j = a_{11}x_1^2 + a_{12}x_1x_2 + \cdots + a_{nn}x_n^2 \tag{6.3.2}$$

over the field $F^{\mathrm{univ}} := \mathbb{Q}(a_{ij})_{i,j=1,\dots,n}$ (now of characteristic zero!) with a_{ij} transcendental elements. We compute its universal determinant

$$\det([T^{\mathrm{univ}}]) = \det\begin{pmatrix} 2a_{11} & a_{12} & \cdots & a_{1n} \\ a_{12} & 2a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{1n} & a_{2n} & \cdots & 2a_{nn} \end{pmatrix} \in 2\mathbb{Z}[a_{ij}]_{i,j} \tag{6.3.3}$$

as a polynomial with integer coefficients. We claim all of these coefficients are even: indeed, reducing modulo 2 and computing the determinant over $\mathbb{F}_2(a_{ij})_{i,j}$, we recall that the determinant of an alternating matrix of odd size is zero (over any field). Therefore, we may let

$$\delta(a_{11},\dots,a_{nn}) := \det(T^{\mathrm{univ}})/2 \in \mathbb{Z}[a_{ij}]_{i,j} \tag{6.3.4}$$

be the **universal (half-)discriminant**. We then define

$$\mathrm{disc}(Q) := \delta(Q(e_1), T(e_1,e_2),\dots,Q(e_n)) \in F/F^{\times 2}$$

by specialization. Repeating the argument in 4.3.2, if $t_{ij}\in F$ and $e_i' := \sum_j t_{ij}e_i$ then

$$\delta(Q(e_1'), T(e_1',e_2'),\dots,Q(e_n')) = \delta(Q(e_1), T(e_1,e_2),\dots,Q(e_n))\det(t_{ij})^2$$

(verified universally!) so $\mathrm{disc}(Q)$ is well-defined. Moreover, this definition agrees with Definition 4.3.3 when $\mathrm{char}\, F \neq 2$.

Example 6.3.5. For example, $\mathrm{disc}(\langle a\rangle) = a$ for $a\in F$, and if

$$Q(x,y,z) = ax^2 + by^2 + cz^2 + uyz + vxz + wxy$$

with $a,b,c,u,v,w\in F$, then

$$\mathrm{disc}(Q) = 4abc + uvw - au^2 - bv^2 - cw^2$$

in all characteristics.

Definition 6.3.6. We say Q is **nondegenerate** if $\mathrm{disc}(Q)\neq 0$.

Next, even though not every quadratic form over F can be diagonalized, so we will also make use of one extra form: for $a,b\in F$, we write $[a,b]$ for the quadratic form $ax^2 + axy + by^2$ on F^2.

Lemma 6.3.7. *There exists a basis of V such that*

$$Q \simeq [a_1, b_1] \boxplus \ldots \boxplus [a_m, b_m] \boxplus \langle c_1, \ldots, c_r \rangle \tag{6.3.8}$$

with $a_i, b_i, c_j \in F$.

Proof. Exercise 6.11. ◻

We say that a quadratic form Q is **normalized** if Q is presented with a basis as in (6.3.8).

Example 6.3.9. The quadratic forms $Q(x, y, z) = x^2 + yz + z^2$ and $Q(x, y, z) = x^2 + y^2 + z^2$ are normalized over $\mathbb{F}_2$, but the quadratic form $Q(x, y, z) = xz + yz + z^2$ is not.

Example 6.3.10. For a normalized quadratic form as in (6.3.8),

$$\begin{aligned}\operatorname{disc}(Q) &= \operatorname{disc}([a_1, b_1] \boxplus \cdots \boxplus [a_m, b_m]) \operatorname{disc}(\langle c_1, \ldots, c_r \rangle) \\ &= (a_1 \cdots a_m)^2 \operatorname{disc}(\langle c_1, \ldots, c_r \rangle).\end{aligned}$$

In $F/F^{\times 2}$, we have

$$\operatorname{disc}(\langle c_1, \ldots, c_r \rangle) = \begin{cases} 0, & \text{if } r \geq 2; \\ c_1, & \text{if } r = 1; \\ 1, & \text{if } r = 0. \end{cases}$$

Therefore, Q is nondegenerate if and only if $a_1 \cdots a_m c_1 \cdots c_r \neq 0$ and $r \leq 1$.

Example 6.3.11. Let $B = \left[\dfrac{a, b}{F}\right)$ be a quaternion algebra. Then $1, i, j, ij$ is a normalized basis for B, and by (6.2.7),

$$\operatorname{nrd} \simeq [1, a] \boxplus [b, ab],$$

so $\operatorname{disc}(\operatorname{nrd}) = b^2$ so nrd is nondegenerate.

6.4 * Characterizing quaternion algebras

We now consider the characterization of quaternion algebras as those equipped with a nondegenerate standard involution (revisiting Main Theorem 4.4.1, but now with $\operatorname{char} F = 2$).

Theorem 6.4.1. *Let B be an F-algebra (with* $\operatorname{char} F = 2$*). Then B has a nondegenerate standard involution if and only if one of the following holds:*

(i) *$B = F$;*
(ii) *$B = K$ is a separable quadratic F-algebra; or*
(iii) *B is a quaternion algebra over F.*

Proof. If $B = F$, then the standard involution is the identity, and nrd is nondegenerate on F because the reduced (half-)discriminant of the quadratic form $\operatorname{nrd}(x) = x^2$ is 1.

If $\dim_F B = 2$, then $B = K$ has a unique standard involution (Lemma 3.4.2). By 6.1.4, we see that the involution is nondegenerate if and only if K is separable.

So suppose $\dim_F B > 2$. Since B has a nondegenerate standard involution, there exists an element $i \in B$ such that $T(i, 1) = \operatorname{trd}(i) \neq 0$. We have $i \notin F$ since $\operatorname{trd}(F) = \{0\}$. Rescaling we may suppose $\operatorname{trd}(i) = 1$, whence $i^2 = i + a$ for some $a \in F$, and $\operatorname{nrd}|_{F+Fi} = [1, a]$. (We have started the proof of Lemma 6.3.7, and $1, i$ is part of a normalized basis, in this special case.)

By nondegeneracy, there exists $j \in \{1, i\}^\perp$ such that $\operatorname{nrd}(j) = b \neq 0$. Thus $\operatorname{trd}(j) = 0$ so $\overline{j} = j$ and $j^2 = b \in F^\times$. Furthermore,

$$0 = \operatorname{trd}(ij) = ij + j\overline{i} = ij + j(i + 1)$$

so $ij = j(i + 1)$. Therefore i, j generate an F-subalgebra $A \simeq [a, b \mid F)$.

The conclusion of the proof follows exactly as in (4.4.3): if $k \in \{1, i, j, ij\}^\perp$ then $k(ij) = k(ji)$, a contradiction. □

Corollary 6.4.2. *Let B be a quaternion algebra over F, and suppose that $K \subseteq B$ is a commutative separable F-subalgebra. Then $\dim_F K \leq 2$. Moreover, if $K \neq F$, then the centralizer of $K^\times$ in $B^\times$ is again $K^\times$.*

Next, we characterize isomorphism classes of quaternion algebras in characteristic 2 in the language of quadratic forms.

6.4.3. Let B be a quaternion algebra over F. We again define

$$B^0 := \{\alpha \in B : \operatorname{trd}(\alpha) = 0\} = \{1\}^\perp. \tag{6.4.4}$$

But now $B^0 = F \oplus Fj \oplus Fk$ and in this basis

$$\operatorname{nrd}(x + yj + zij) = x^2 + by^2 + byz + abz^2 \tag{6.4.5}$$

so $\operatorname{nrd}|_{B^0} \simeq \langle 1 \rangle \boxplus [b, ab]$. The discriminant is therefore

$$\operatorname{disc}(\operatorname{nrd}|_{B^0}) = b^2 = 1 \in F^\times/F^{\times 2}. \tag{6.4.6}$$

Theorem 6.4.7. *Let F be a field with $\operatorname{char} F = 2$. Then the functor $B \mapsto \operatorname{nrd}|_{B^0}$ yields an equivalence of categories between*

Quaternion algebras over F,
under F-algebra isomorphisms

and

Ternary quadratic forms over F with discriminant $1 \in F^\times/F^{\times 2}$,
under isometries.

Proof. We argue as in Theorem 5.6.8 but with $\operatorname{char} F = 2$. The argument here is easier, because all sign issues go away and there is no orientation to chase: by Exercise 6.12, there is a unique $\zeta \in \operatorname{Clf}^1 Q \smallsetminus F$ such that $\zeta^2 = 1$. The inclusion $\iota\colon B^0 \hookrightarrow B$ induces a surjective F-algebra homomorphism $\operatorname{Clf}^0(\operatorname{nrd}|_{B^0}) \to B$, so by dimensions it is an isomorphism; this gives one natural transformation. In the other direction, the map $m_\zeta : V \to B^0$ by $v \mapsto v\zeta$ is again an isometry by (5.6.10), giving the other.

Here is a second direct proof. By 6.4.3, the quadratic form $\operatorname{nrd}|_{B^0}$ has discriminant 1. To show the functor is essentially surjective, let $Q\colon V \to F$ be a ternary quadratic form with discriminant $1 \in F^\times/F^{\times 2}$. Then $Q \simeq \langle u\rangle \boxplus [b,c]$ for some $u,b,c \in F$. We have $\operatorname{disc}(Q) = ub^2 = 1 \in F^{\times 2}$ so $b \in F^\times$ and $u \in F^{\times 2}$. Rescaling the first variable, we obtain $Q \sim \langle 1\rangle \boxplus [b,c]$. Thus by 6.4.3, Q arises up to isometry from the quaternion algebra $\left[\dfrac{a,b}{F}\right)$ with $a = cb^{-1}$.

For morphisms, we argue as in the proof of Proposition 5.2.4 but with $\operatorname{char} F = 2$. In one direction, an F-algebra isomorphism $B \xrightarrow{\sim} B'$ induces an isometry $B^0 \xrightarrow{\sim} (B')^0$ by uniqueness of the standard involution. Conversely, let $f\colon B^0 \to (B')^0$ be an isometry. Let $B \simeq \left[\dfrac{a,b}{F}\right)$. Extend f to an F-linear map $B \to B'$ by mapping $i \mapsto b^{-1}f(ij)f(j)$. The map f preserves 1: it maps F to F by Exercise 6.15, since $F = (B^0)^\perp = ((B')^0)^\perp$, and $1 = \operatorname{nrd}(1) = \operatorname{nrd}(f(1)) = f(1)^2$ so $f(1) = 1$. We have $f(j)^2 = \operatorname{nrd}(f(j)) = \operatorname{nrd}(j) = b$ and similarly $f(ij)^2 = ab$ since $j, ij \in B^0$. Thus

$$\begin{aligned} 1 = \operatorname{trd}(i) &= b^{-1}\operatorname{trd}((ij)j) = b^{-1}T(ij,j) = \\ &= b^{-1}T(f(ij),f(j)) = \operatorname{trd}(b^{-1}f(ij)f(j)) = \operatorname{trd}(f(i)) \end{aligned}$$

and similarly $\operatorname{nrd}(f(i)) = \operatorname{nrd}(i) = a$, thus $f(i)^2 + f(i) + a = 0$. Finally,

$$f(i)f(j) = b^{-1}f(ij)f(j)^2 = f(ij)$$

and

$$\begin{aligned} f(j)f(i) &= b^{-1}f(j)f(ij)f(j) = b^{-1}f(j)(f(j)f(ij) + T(f(j),f(ij))) \\ &= f(ij) + f(j) = (f(i)+1)f(j) = \overline{f(i)}f(j) \end{aligned}$$

so f is an isomorphism of F-algebras. Therefore the functor is full and faithful, yielding an equivalence of categories. □

Corollary 6.4.8. *The maps $B \mapsto Q = \operatorname{nrd}|_{B^0} \mapsto C = V(Q)$ yield bijections*

$$\left\{\begin{array}{c}\text{Quaternion algebras over } F\\ \text{up to isomorphism}\end{array}\right\} \leftrightarrow \left\{\begin{array}{c}\text{Nondegenerate ternary}\\ \text{quadratic forms over } F\\ \text{with discriminant } 1 \in F^\times/F^{\times 2}\\ \text{up to isometry}\end{array}\right\}$$

$$\leftrightarrow \left\{\begin{array}{c}\text{Nondegenerate ternary}\\ \text{quadratic forms over } F\\ \text{up to similarity}\end{array}\right\}$$

$$\leftrightarrow \left\{\begin{array}{c}\text{Conics over } F\\ \text{up to isomorphism}\end{array}\right\}$$

that are functorial with respect to F.

Proof. The remaining parts of the bijection follow as in the proof of Corollary 5.2.6. ◻

We now turn to identifying the matrix ring in characteristic 2.

Definition 6.4.9. A quadratic form $H\colon V \to F$ is a **hyperbolic plane** if $H \simeq [1,0]$.

Recall that $[1,0] : F^2 \to F$ is given by the quadratic form $x^2 + xy = x(x+y)$, visibly isometric to the quadratic form xy. Definition 6.4.9 agrees with Definition 5.4.1 after a change of basis.

Lemma 6.4.10. *If Q is nondegenerate and isotropic then $Q \simeq H \boxplus Q'$ with H a hyperbolic plane.*

Proof. We repeat the proof of Lemma 5.4.2. ◻

We may again characterize division quaternion algebras by examination of the reduced norm as a quadratic form as in Main Theorem 5.4.4 and Theorem 5.5.3.

Theorem 6.4.11. *Let $B = \left[\dfrac{a,b}{F}\right)$ (with $\operatorname{char} F = 2$). Then the following are equivalent:*

(i) $B \simeq \left[\dfrac{1,1}{F}\right) \simeq \mathrm{M}_2(F)$;
(ii) *B is not a division ring;*
(iii) *The quadratic form nrd is isotropic;*
(iv) *The quadratic form $\operatorname{nrd}|_{B^0}$ is isotropic;*
(v) *The binary form $[1,a]$ represents b;*
(vi) *$b \in \operatorname{Nm}_{K|F}(K^\times)$ where $K = F[i]$; and*
(vii) *The conic $C := V(\operatorname{nrd}|_{B^0}) \subset \mathbb{P}^2$ has an F-rational point.*

Proof. Only condition (v) requires significant modification in the case $\operatorname{char} F = 2$; see Exercise 6.13. ◻

Lemma 6.4.12. *Let $K \supset F$ be a quadratic extension of fields. Then K is a splitting field for B if and only if there is an injective F-algebra homomorphism $K \hookrightarrow B$.*

Proof. If $\iota\colon K \hookrightarrow B$ and $K = F(\alpha)$, then $1 \otimes \alpha - \iota(\alpha) \otimes 1$ is a zerodivisor in $B \otimes_F K$, since

$$(1 \otimes \alpha - \iota(\alpha) \otimes 1)\,(1 \otimes \alpha - \iota(\overline{\alpha}) \otimes 1) = 0, \tag{6.4.13}$$

and so $B \otimes_F K \simeq \mathrm{M}_2(K)$ and K is a splitting field.

Conversely, let $K = F(\alpha)$ and suppose $B \otimes_F K \simeq \mathrm{M}_2(K)$. If $B \simeq \mathrm{M}_2(F)$, we can take the embedding mapping α to a matrix with the same rational canonical form. So we suppose that $B = \left[\dfrac{a,b}{F}\right)$ is a division ring. By Theorem 6.4.11(iv) (over K) and 6.4.3, there exist $x,y,z,u,v,w \in F$ not all zero such that

$$(x+u\alpha)^2 + b(y+v\alpha)^2 + b(y+v\alpha)(z+w\alpha) + ab(z+w\alpha)^2 = 0; \tag{6.4.14}$$

expanding and rewriting into the powers of α gives

$$(u^2 + bv^2 + bvw + abw^2)\alpha^2 + (vz + wy)b\alpha + (x^2 + by^2 + byz + abz^2) = 0. \quad (6.4.15)$$

Let $\beta := x + yj + zij$ and $\gamma := u + vj + wij$. Then $\gamma \in B^\times$, since $\gamma = 0$ implies $\operatorname{nrd}(\beta) = 0$ and yet B is a division ring. Then the equation (6.4.15) can be written

$$\operatorname{nrd}(\gamma)\alpha^2 + \operatorname{trd}(\beta\gamma)\alpha + \operatorname{nrd}(\beta) = 0;$$

then a direct calculation shows that the element

$$\mu = \beta\gamma^{-1} = \operatorname{nrd}(\gamma)^{-1}\beta\gamma$$

satisfies the same equation as (6.4.15) in the variable α, so there is an embedding $K \hookrightarrow B$ defined by $\alpha \mapsto \mu$. □

Exercises

Throughout these exercises, we let F be a field (of any characteristic, unless specified).

1. Recall the primitive element theorem from Galois theory: if $K \supseteq F$ is a separable field extension of finite degree, then there exists $\alpha \in K$ such that $K = F(\alpha)$—and hence $K \simeq F[x]/(f(x))$ where $f(x) \in F[x]$ is the minimal polynomial of α. Extend this theorem to algebras as follows. Let B be a separable, commutative, finite-dimensional F-algebra. Show that $B \simeq F[x]/(f(x))$ for some $f(x) \in F[x]$.
▷ 2. Let B be a quaternion algebra over F and let $K \subset B$ be a separable quadratic F-algebra. Show that there exists $b \in F^\times$ such that $B \simeq \left(\dfrac{K, b}{F}\right)$ (as in 6.1.5).
3. Let F^{sep} be a separable closure of F and let B be a quaternion algebra over F. Show that $B \otimes_F F^{\mathrm{sep}} \simeq \mathrm{M}_2(F^{\mathrm{sep}})$. [More generally, see Exercise 7.24.]
▷ 4. Let K be a separable quadratic F-algebra and let $u, b \in F^\times$. Show that $\left(\dfrac{K, b}{F}\right) \simeq \left(\dfrac{K, ub}{F}\right)$ if and only if $u \in \operatorname{nrd}(K^\times) = \operatorname{Nm}_{K|F}(K^\times)$.
5. Let B be a quaternion algebra over F, and let $K_0 \supseteq F$ be a quadratic field. Prove that there exists a separable extension $K \supseteq F$ linearly disjoint from K_0 over F (i.e., $K \otimes_F K_0$ is a domain) such that K splits F.
6. Suppose $\operatorname{char} F = 2$ and let $a \in F$ and $b \in F^\times$.
   (a) Show that $\left[\dfrac{a, b}{F}\right) \simeq \left[\dfrac{a, ab}{F}\right)$ if $a \neq 0$.
   (b) Show that if $t \in F$ and $u \in F^\times$, then $\left[\dfrac{a, b}{F}\right) \simeq \left[\dfrac{a + (t + t^2), bu^2}{F}\right)$.
7. Let $\operatorname{char} F = 2$ and let $B = \left[\dfrac{a, b}{F}\right)$ be a quaternion algebra over F. Compute the left regular representation $\lambda \colon B \to \operatorname{End}_K(B) \simeq \mathrm{M}_2(K)$ where $K = F[i]$ as in 2.3.8.

▹ 8. Suppose $\operatorname{char} F = 2$. Let $M \in \mathrm{M}_n(F)$ be a symmetric matrix with n odd, and suppose that all diagonal entries of M are zero. Show that $\det M = 0$.

▹ 9. Let $\operatorname{char} F = 2$ and let B be a division F-algebra with a standard involution. Prove that either the standard involution is the identity (and so B is classified by Exercise 3.9), or that the conclusion of Theorem 3.5.1 holds for B: namely, that either $B = K$ is a separable quadratic field extension of F or that B is a quaternion algebra over F. *[Hint: Replace conjugation by i by the map $\phi(x) = ix + xi$, and show that $\phi^2 = \phi$. Then diagonalize and proceed as in the case* $\operatorname{char} F \neq 2$.*]*

▹ 10. Let $\operatorname{char} F = 2$. Show that the even Clifford algebra $\operatorname{Clf}^0 Q$ of a nondegenerate ternary quadratic form $Q\colon V \to F$ is a quaternion algebra over F.

▹ 11. Prove Lemma 6.3.7, that every quadratic form over F with $\operatorname{char} F = 2$ has a normalized basis.

▹ 12. Let $\operatorname{char} F = 2$ and let $Q\colon V \to F$ be a quadratic form over F with discriminant $d \in F^\times/F^{\times 2}$ and $\dim_F V = n$ odd. Show that $Z(\operatorname{Clf} Q) \simeq F[x]/(x^2 - d)$ and that there is a unique $\zeta \in Z(\operatorname{Clf} Q) \cap \operatorname{Clf}^1 Q$ such that $\zeta^2 = 1$.

▹ 13. Prove Theorem 6.4.11.

▹ 14. Let $Q := Q' \boxplus Q''$ be an orthogonal sum of two anisotropic quadratic forms over F (with F of arbitrary characteristic). Show that Q is isotropic if and only if there exists $c \in F^\times$ that is represented by both Q' and $-Q''$.

▹ 15. Let B be a quaternion algebra over F (with F of arbitrary characteristic). Show that $F = (B^0)^\perp$.

▹ 16. Prove Wedderburn's little theorem in the following special case: a quaternion algebra over a finite field with even cardinality is not a division ring. *[Hint: See Exercise 3.16.]*

This chapter is licensed under the terms of the Creative Commons Attribution-NonCommercial 4.0 International License (http://creativecommons.org/licenses/by-nc/4.0/), which permits any noncommercial use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license and indicate if changes were made.

The images or other third party material in this chapter are included in the chapter's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the chapter's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.

Chapter 7

Simple algebras

In this chapter, we return to the characterization of quaternion algebras. We initially defined quaternion algebras in terms of generators and relations in Chapter 2; in the chapters that followed, we showed that quaternion algebras are equivalently noncommutative algebras with a nondegenerate standard involution. Here, we pursue another approach, and we characterize quaternion algebras in a different way, as central simple algebras of dimension 4.

7.1 ▹ Motivation and summary

Consider now the "simplest" sorts of algebras. Like the primes among the integers or the finite simple groups among finite groups, it is natural to seek algebras that cannot be "broken down" any further. Accordingly, we say that a ring A is **simple** if it has no nontrivial two-sided (bilateral) ideals, i.e., the only two-sided ideals are $\{0\}$ and A. To show the power of this notion, consider this: if $\phi\colon A \to A'$ is a ring homomorphism and A is simple, then ϕ is either injective or the zero map (since $\ker \phi \subseteq B$ is a two-sided ideal).

A division ring A is simple, since every nonzero element is a unit and therefore every nonzero ideal (left, right, or two-sided) contains 1 so is equal to A. In particular, a field is a simple ring, and a *commutative* ring is simple if and only if it is a field. The matrix ring $\mathrm{M}_n(F)$ over a field F is also simple, something that can be checked directly by multiplying by matrix units (Exercise 7.5).

Moreover, quaternion algebras are simple. The shortest proof of this statement, given what we have done so far, is to employ Main Theorem 5.4.4 (and Theorem 6.4.11 in characteristic 2): a quaternion algebra B over F is either isomorphic to $\mathrm{M}_2(F)$ or is a division ring, and in either case is simple. One can also prove this directly (Exercise 7.1).

Although the primes are quite mysterious and the classification of finite simple groups is a monumental achievement in group theory, the situation for algebras is quite simple, indeed! Our first main result is as follows (Main Theorem 7.3.10).

© The Author(s) 2021

J. Voight, *Quaternion Algebras*, Graduate Texts in Mathematics 288,
https://doi.org/10.1007/978-3-030-56694-4_7

Theorem 7.1.1. (Wedderburn–Artin). *Let F be a field and B be a finite-dimensional F-algebra. Then B is simple if and only if $B \simeq \mathrm{M}_n(D)$ where $n \geq 1$ and D is a finite-dimensional division F-algebra.*

As a corollary of Theorem 7.1.1, we give another characterization of quaternion algebras.

Corollary 7.1.2. *Let B be an F-algebra. Then the following are equivalent:*

(i) *B is a quaternion algebra;*
(ii) *$B \otimes_F F^{\mathrm{al}} \simeq \mathrm{M}_2(F^{\mathrm{al}})$, where F^{al} is an algebraic closure of F; and*
(iii) *B is a central simple algebra of dimension $\dim_F B = 4$.*

Moreover, a central simple algebra B of dimension $\dim_F B = 4$ is either a division algebra or has $B \simeq \mathrm{M}_2(F)$.

This corollary has the neat consequence that a division algebra B over F is a quaternion algebra over F if and only if it is central of dimension $\dim_F B = 4$.

For the reader in a hurry, we now give a proof of this corollary without invoking the Wedderburn–Artin theorem; this proof also serves as a preview of some of the ideas that go into the theorem.

Proof of Corollary 7.1.2. The statement (i) $\Rightarrow$ (ii) was proven in Exercise 2.4(d).

To prove (ii) $\Rightarrow$ (iii), suppose B is an algebra with $B^{\mathrm{al}} := B \otimes_F F^{\mathrm{al}} \simeq \mathrm{M}_2(F^{\mathrm{al}})$. The F^{al}-algebra B^{al} is central simple, from above. Thus $Z(B) = Z(B^{\mathrm{al}}) \cap B = F$. And if I is a two-sided ideal of B then $I^{\mathrm{al}} := I \otimes_F F^{\mathrm{al}}$ is a two-sided ideal of B^{al}, so $I^{\mathrm{al}} = \{0\}$ or $I^{\mathrm{al}} = B^{\mathrm{al}}$ is trivial, whence $I = I^{\mathrm{al}} \cap F$ is trivial. Finally, $\dim_F B = \dim_{F^{\mathrm{al}}} B^{\mathrm{al}} = 4$.

Finally, we prove (iii) $\Rightarrow$ (i). Let B a central simple F-algebra of dimension 4. If B is a division algebra we are done; so suppose not. Then B has a nontrivial left ideal (e.g., one generated by a nonunit); let $\{0\} \subsetneq I \subsetneq B$ be a nontrivial left ideal with $0 < m = \dim_F I$ minimal. Then there is a nonzero homomorphism $B \to \mathrm{End}_F(I) \simeq \mathrm{M}_m(F)$ which is injective, since B is simple. By dimension, we cannot have $m = 1$; if $m = 2$, then $B \simeq \mathrm{M}_2(F)$ and we are done. So suppose $m = 3$. Then by minimality, every nontrivial left ideal of B has dimension 3. But for any $\alpha \in B$, we have that $I\alpha$ is a left ideal, so the left ideal $I \cap I\alpha$ is either $\{0\}$ or I; in either case, $I\alpha \subseteq I$ and I is a right ideal as well. But this contradicts the fact that B is simple. □

The Wedderburn–Artin theorem is an important structural result used throughout mathematics, so we give in this chapter a self-contained account of its proof. More generally, it will be convenient to work with **semisimple** algebras, finite direct products of simple algebras. When treating ideals of an algebra we would be remiss if we did not discuss more generally modules over the algebra, and the notions of simple and semisimple module are natural concepts in linear algebra and representation theory: a semisimple module is one that is a direct sum of simple modules ("completely reducible"), analogous to a semisimple operator where every invariant subspace has an invariant complement (e.g., a diagonalizable matrix).

The second important result in this chapter is a theorem that concerns the simple subalgebras of a simple algebra, as follows (Main Theorem 7.7.1).

Theorem 7.1.3. (Skolem–Noether). *Let A, B be simple F-algebras and suppose that B is central. Suppose that $f, g\colon A \to B$ are homomorphisms. Then there exists $\beta \in B$ such that $f(\alpha) = \beta^{-1} g(\alpha)\beta$ for all $\alpha \in A$.*

Corollary 7.1.4. *Every F-algebra automorphism of a simple F-algebra B is inner, i.e.,* $\operatorname{Aut}(B) \simeq B^\times/F^\times$.

Just as above, for our quaternionic purposes, we can give a direct proof.

Corollary 7.1.5. *Let B be a quaternion algebra over F and let $K_1, K_2 \subset B$ be quadratic subfields. Suppose that $\phi\colon K_1 \xrightarrow{\sim} K_2$ is an isomorphism of F-algebras. Then ϕ lifts to an inner automorphism of B, i.e., there exists $\beta \in B$ such that $\alpha_2 = \phi(\alpha_1) = \beta^{-1}\alpha_1\beta$ for all $\alpha_1 \in K_1$. In particular, $K_2 = \beta^{-1}K_1\beta$.*

Proof. Write $K_1 = F(\alpha_1)$ with $\alpha_1 \in B$ and let $\alpha_2 = \phi(\alpha_1) \in K_2 \subset B$, so $K_2 = F(\alpha_2)$. We want to find $\beta \in B^\times$ such that $\alpha_2 = \beta^{-1}\alpha_1\beta$. In the special case $B \simeq \mathrm{M}_2(F)$, then $\alpha_1, \alpha_2 \in \mathrm{M}_2(F)$ satisfy the same *irreducible* characteristic polynomial, so by the theory of rational canonical forms, $\alpha_2 = \beta^{-1}\alpha_1\beta$ where $\beta \in B^\times \simeq \mathrm{GL}_2(F)$ as desired.

Suppose then that B is a division ring. Then the set

$$W = \{\beta \in B : \beta\alpha_2 = \alpha_1\beta\} \tag{7.1.6}$$

is an F-vector subspace of B. Let F^{sep} be a separable closure of F. (Or, apply Exercise 6.5 and work over a splitting field K linearly disjoint from $K_1 \simeq K_2$.) Then we have $B \otimes_F F^{\mathrm{sep}} \simeq \mathrm{M}_2(F^{\mathrm{sep}})$, and the common characteristic polynomial of α_1, α_2 either remains irreducible over F^{sep} (if $K \supset F$ is inseparable) or splits as the product of two linear factors with distinct roots. In either case, the theory of rational canonical forms again applies, and there exists $\beta \in B \otimes_F F^{\mathrm{sep}} \simeq \mathrm{GL}_2(F^{\mathrm{sep}})$ that will do; but then by linear algebra $\dim_{F^{\mathrm{sep}}} W \otimes_F F^{\mathrm{sep}} = \dim_F W > 0$, so there exists $\beta \in B \setminus \{0\} = B^\times$ with the desired property. $\square$

As shown in the above proof, Corollary 7.1.5 can be seen as a general reformulation of the rational canonical form from linear algebra.

7.2 Simple modules

Basic references for this section include Drozd–Kirichenko [DK94, §1–4], Curtis–Reiner [CR81, §3], Lam [Lam2001, §2–3], and Farb–Dennis [FD93, Part I]. An elementary approach to the Weddernburn–Artin theorem is given by Brešar [Bre2010]. An overview of the subject of associative algebras is given by Pierce [Pie82] and Jacobson [Jacn2009].

Throughout this chapter, let B be a finite-dimensional F-algebra.

To understand the algebra B, we look at its representations. A **representation** of B (over F) is a vector space V over F together with an F-algebra homomorphism $B \to \operatorname{End}_F(V)$. Equivalently, a representation is given by a left (or right) B-module V: this is almost a tautology. Although one can define infinite-dimensional representations, they will not interest us here, and we suppose throughout that $\dim_F V < \infty$, or

equivalently that V is a finitely generated (left or right) B-module. If we choose a basis for V, we obtain an isomorphism $\operatorname{End}_F(V) \simeq \mathrm{M}_n(F)$ where $n = \dim_F V$, so a representation is just a homomorphic way of thinking of the algebra B as an algebra of matrices.

Example 7.2.1. The space of column vectors F^n is a left $\mathrm{M}_n(F)$-module; the space of row vectors is a right $\mathrm{M}_n(F)$-module.

Example 7.2.2. B is itself a left B-module, giving rise to the **left regular representation** $B \to \operatorname{End}_F(B)$ over F (cf. Remark 3.3.8).

Example 7.2.3. Let G be a finite group. Then a representation of $F[G]$ (is the same as an $F[G]$-module which) is the same as a homomorphism $G \to \mathrm{GL}(V)$, where V is an F-vector space (Exercise 3.8).

Definition 7.2.4. Let V be a left B-module. We say V is **simple** (or **irreducible**) if $V \neq \{0\}$ and the only B-submodules of V are $\{0\}$ and V.

We say V is **indecomposable** if V cannot be written as $V = V_1 \oplus V_2$ with $V_1, V_2 \neq \{0\}$ left B-modules.

A simple module is indecomposable, but the converse need not hold, and this is a central point of difficulty in understanding representations.

Example 7.2.5. If $B = \left\{\begin{pmatrix} a & b \\ 0 & c \end{pmatrix} : a, b, c \in F\right\} \subseteq \mathrm{M}_2(F)$, then the space $V = F^2$ of column vectors is not simple, since the subspace spanned by $\begin{pmatrix} 1 \\ 0 \end{pmatrix}$ is a B-submodule; nevertheless, V is indecomposable (Exercise 7.4).

The importance of simple modules is analogous to that of simple groups. Arguing by induction on the dimension of V, we have the following lemma analogous to the Jordan–Hölder theorem on composition series.

Lemma 7.2.6. *A (finite-dimensional) left B-module V admits a filtration*

$$V = V_0 \supsetneq V_1 \supsetneq V_2 \supsetneq \cdots \supsetneq V_r = \{0\}$$

such that V_i/V_{i+1} is simple for each i.

This filtration is not unique, but up to isomorphism and permutation, the quotients V_i/V_{i+1} are unique.

Lemma 7.2.7. *If I is a maximal left ideal of B, then B/I is a simple B-module. Conversely, if V is a simple B-module, then $V \simeq B/I$ for a maximal left ideal I: more precisely, for any $x \in V \smallsetminus \{0\}$, we may take*

$$I = \operatorname{ann}(x) := \{\alpha \in B : \alpha x = 0\}.$$

Proof. For the first statement, a submodule of B/I corresponds to a left ideal containing I, so B/I is simple if and only if I is maximal. Conversely, letting $x \in V \smallsetminus \{0\}$ we have $\{0\} \neq Bx \subseteq V$ a B-submodule and so $Bx = V$; and consequently $V \simeq B/I$ where $I = \operatorname{ann}(x)$ and again I is a maximal left ideal. ∎

Having defined the notion of simplicity for modules, we now consider simplicity of the algebra B.

Definition 7.2.8. An F-algebra B is **simple** if the only two-sided ideals of B are $\{0\}$ and B.

Equivalently, B is simple if and only if any F-algebra (or even ring) homomorphism $B \to A$ is either injective or the zero map.

Example 7.2.9. A division F-algebra D is simple. In fact, the F-algebra $\mathrm{M}_n(D)$ is simple for any division F-algebra D (Exercise 7.5), and in particular $\mathrm{M}_n(F)$ is simple.

Example 7.2.10. Let F^{al} be an algebraic closure of F. If $B \otimes_F F^{\mathrm{al}}$ is simple, then B is simple. The association $I \mapsto I \otimes_F F^{\mathrm{al}}$ is an injective map from the set of two-sided ideals of B to the set of two-sided ideals of $B \otimes_F F^{\mathrm{al}}$.

7.2.11. If B is a quaternion algebra over F, then B is simple. We have $B \otimes_F F^{\mathrm{al}} \simeq \mathrm{M}_2(F^{\mathrm{al}})$, which is simple by Example 7.2.9, and B is simple by Example 7.2.10.

Example 7.2.9 shows that algebras of the form $\mathrm{M}_n(D)$ with D a division F-algebra yield a large class of simple F-algebras. In fact, these are all such algebras, a fact we will now prove. First, a few preliminary results.

Lemma 7.2.12. (Schur). *Let B be an F-algebra. Let V_1, V_2 be simple B-modules. Then any homomorphism $\phi\colon V_1 \to V_2$ of B-modules is either zero or an isomorphism.*

Proof. We have that $\ker\phi$ and $\operatorname{img}\phi$ are B-submodules of V_1 and V_2, respectively, so either $\phi = 0$ or $\ker\phi = \{0\}$ and $\operatorname{img}\phi = V_2$, hence $V_1 \simeq V_2$. ◻

Corollary 7.2.13. *If V is a simple B-module, then* $\mathrm{End}_B(V)$ *is a division ring.*

7.2.14. Let B be an F-algebra and consider B as a left B-module. Then there is a map

$$\begin{aligned} \rho\colon B^{\mathrm{op}} &\to \mathrm{End}_B(B) \\ \alpha &\mapsto (\rho_\alpha : \beta \mapsto \beta\alpha), \end{aligned}$$

where B^{op} is the opposite algebra of B defined in 3.2.2. The map ρ is injective since $\rho_\alpha = 0$ implies $\rho_\alpha(1) = \alpha = 0$; it is also surjective, since if $\phi \in \mathrm{End}_B(B)$ then letting $\alpha = \phi(1)$ we have $\phi(\beta) = \beta\phi(1) = \beta\alpha$ for all $\beta \in B$. Finally, it is an F-algebra homomorphism, since

$$\rho_{\alpha\beta}(\mu) = \mu(\alpha\beta) = (\mu\alpha)\beta = (\rho_\beta \circ \rho_\alpha)(\mu),$$

and therefore ρ is an isomorphism of F-algebras.

One lesson here is that a left module has endomorphisms that act naturally on the right; but the more common convention is that endomorphisms also act on the left. In order to make this compatible, the opposite algebra intervenes.

7.2.15. More generally, the decomposition of modules is determined by idempotent endomorphisms as follows. Let V be a left B-module. Then V is indecomposable if and only if $\operatorname{End}_B(V)$ has no nontrivial **idempotents**: that is to say, if $e \in \operatorname{End}_B(V)$ satisfies $e^2 = e$, then $e \in \{0, 1\}$. Given a nontrivial idempotent, we can write $V = eV \oplus (1 - e)V$, and conversely if $V = V_1 \oplus V_2$ then the projection $V \to V_1 \subseteq V$ gives an idempotent.

7.2.16. Many theorems of linear algebra hold equally well over division rings as they do over fields, as long as one is careful about the direction of scalar multiplication. For example, let D be a division F-algebra and let V be a left D-module. Then $V \simeq D^n$ is free, and choice of basis for V gives an isomorphism $\operatorname{End}_D(V) \simeq \mathrm{M}_n(D^{\mathrm{op}})$. When $n = 1$, this becomes $\operatorname{End}_D(D) \simeq D^{\mathrm{op}}$, as in 7.2.14.

Lemma 7.2.17. *Let B be a (finite-dimensional) simple F-algebra. Then there exists a simple left B-module which is unique up to isomorphism.*

Proof. Since B is finite-dimensional over F, there is a nonzero left ideal I of B of minimal dimension, and such an ideal I is necessarily simple. Moreover, if $v \in I$ is nonzero then $Bv = I$, since $Bv \subseteq I$ is nonzero and I is simple. Let $I = Bv$ with $v \in I$.

Now let V be any simple B-module; we will show $I \simeq V$ as B-modules. Since B is simple, the natural map $B \to \operatorname{End}_F(V)$ is injective (since it is nonzero). Therefore, there exists $x \in V$ such that $vx \neq 0$, so $Ix \neq \{0\}$. Thus, the map $I \to V$ by $\beta \mapsto \beta x$ is a nonzero B-module homomorphism, so it is an isomorphism by Schur's lemma. □

Example 7.2.18. The unique simple left $\mathrm{M}_n(F)$-module (up to isomorphism) is the space F^n of column vectors (Example 7.2.1).

7.2.19. Every algebra can be decomposed according to its idempotents 7.2.15. Let B be a finite-dimensional F-algebra. Then we can write $B = I_1 \oplus \cdots \oplus I_r$ as a direct sum of *indecomposable* left B-modules: this follows by induction, as the decomposing procedure must stop because each factor is a finite-dimensional F-vector space. This means we may write

$$1 = e_1 + \cdots + e_r$$

with $e_i \in I_i$. For each $\alpha \in I_i$ we have $\alpha = \sum_i \alpha e_i$ whence $\alpha e_i = \alpha$ and $\alpha e_j = 0$ for $j \neq i$, which implies that

$$e_i^2 = e_i, \quad e_i e_j = 0 \quad \text{for } j \neq i, \qquad \text{and } I_i = Be_i.$$

Thus each e_i is idempotent; we call $\{e_1, \dots, e_r\}$ a complete set of **primitive orthogonal idempotents**: the *orthogonal* is because $e_i e_j = 0$ for $j \neq i$, and the *primitive* is because each e_i is not the sum of two other orthogonal idempotents (by 7.2.15).

Remark 7.2.20. The tight connection between F and $\mathrm{M}_n(F)$ is encoded in the fact that the two rings are **Morita equivalent**: there is an equivalence of categories between F-vector spaces and left $\mathrm{M}_n(F)$-modules. For more on this rich subject, see Lam [Lam99, §18], Reiner [Rei2003, Chapter 4], and Curtis–Reiner [CR81, §35].

7.3 Semisimple modules and the Wedderburn–Artin theorem

We continue our assumptions that B is a finite-dimensional F-algebra and a B-module V is finite-dimensional.

Definition 7.3.1. A B-module V is **semisimple** (or **completely reducible**) if V is isomorphic to a (finite) direct sum of simple B-modules $V \simeq \bigoplus_i V_i$.

B is a **semisimple** F-algebra if B is semisimple as a left B-module.

Remark 7.3.2. More precisely, we have defined the notion of **left semisimple** and could equally well define **right semisimple**; below we will see that these two notions are the same.

Example 7.3.3. If $B = F$, then simple F-modules are one-dimensional vector spaces, and as F is simple these are the only ones. Every F-vector space has a basis and so is the direct sum of one-dimensional subspaces, thus every F-module is semisimple.

Example 7.3.4. A finite-dimensional commutative F-algebra B is semisimple if and only if B is the product of field extensions of F, i.e., $B \simeq K_1 \times \cdots \times K_r$ with $K_i \supseteq F$ a finite extension of fields.

Lemma 7.3.5. *The following statements hold.*

(a) *A B-module V is semisimple if and only if it is the sum of simple B-modules.*
(b) *A submodule or a quotient module of a semisimple B-module is semisimple.*
(c) *If B is a semisimple F-algebra, then every B-module is semisimple.*

Proof. For (a), let $V = \sum_i V_i$ be the sum of simple B-modules. Since V is finite-dimensional, we can rewrite it as an irredundant finite sum; and then since each V_i is simple, the intersection of any two distinct summands is $\{0\}$, so the sum is direct.

For (b), let $W \subseteq V$ be a submodule of the semisimple B-module V. Every $x \in W$ with $x \neq 0$ is contained in a simple B-submodule of W by minimality, so $W = \sum_i W_i$ is a sum of simple B-modules. The result now follows from (a) for submodules. For quotient modules, suppose $\phi\colon V \to Z$ is a surjective B-module homomorphism; then $\phi^{-1}(Z) \subseteq V$ is a B-submodule, and $\phi^{-1}(Z) = \sum_i W_i$ is a sum of simple B-modules, and hence by Schur's lemma $Z = \sum_i \phi(W_i)$ is semisimple.

For (c), let V be a B-module. Since V is finitely generated as a B-module, there is a surjective B-module homomorphism $B^r \to V$ for some $r \geq 1$. Since B^r is semisimple, so too is V by (b). □

Remark 7.3.6. Doing linear algebra with semisimple modules mirrors very closely linear algebra over a field. We have already seen that every submodule and quotient module of a semisimple module is again semisimple. Moreover, every module homomorphism $V \to W$ with V semisimple splits, and every submodule of a semisimple module is a direct summand. The extent to which this fails over other rings concerns the structure of projective modules; we take this up in Chapter 20.

Lemma 7.3.7. *If B is a simple F-algebra, then B is a semisimple F-algebra.*

Proof. Let $I \subseteq B$ be a minimal nonzero left ideal, the unique simple left B-module up to isomorphism as in Lemma 7.2.17. For all $\alpha \in B$, the left ideal $I\alpha$ is a homomorphic image of I, so by Schur's lemma, either $I\alpha = \{0\}$ or $I\alpha$ is simple. Let $A := \sum_{\alpha \in B} I\,\alpha$. Then A is a nonzero two-sided ideal of B, so since B is simple, we conclude $A = B$. Thus B is the sum of simple B-modules, and the result follows from Lemma 7.3.5(a). □

Corollary 7.3.8. *A (finite) direct product of simple F-algebras is a semisimple F-algebra.*

Proof. If $B \simeq B_1 \times \cdots \times B_r$ with each B_i simple, then by Lemma 7.3.7, each B_i is semisimple so $B_i = \bigoplus_j I_{ij}$ is the direct sum of simple B_i-modules I_{ij}. Each I_{ij} has the natural structure of a B-module (extending by zero), and with this structure it is simple, and $B = \bigoplus_{i,j} I_{ij}$ is semisimple. □

The converse of Corollary 7.3.8 is true and is proven as Corollary 7.3.14, a consequence of the Wedderburn–Artin theorem.

In analogy to 7.2.16, we have the following corollary.

Corollary 7.3.9. *Let B be a simple F-algebra and let V be a left B-module. Then $V \simeq I^{\oplus n}$ for some $n \geq 1$, where I is a simple left B-module. In particular, two left B-modules V_1, V_2 are isomorphic if and only if $\dim_F V_1 = \dim_F V_2$.*

Proof. Since B is simple, B is semisimple by Lemma 7.3.7, and V is semisimple by Lemma 7.3.5. But by Lemma 7.2.17, there is a unique simple left B-module I, and the result follows. □

In other words, this corollary says that if B is simple then every left B-module V is free over B, so has a left basis over B; if we define the **rank** of a left B-module V to be cardinality of this basis (the integer n such that $V \simeq I^{\oplus n}$ as in Corollary 7.3.9), then two such modules are isomorphic if and only if they have the same rank.

We now come to one of the main results of this chapter.

Main Theorem 7.3.10. (Wedderburn–Artin). *Let B be a finite-dimensional F-algebra. Then B is semisimple if and only if there exist integers $n_1, \dots, n_r$ and division algebras $D_1, \dots, D_r$ such that*

$$B \simeq \mathrm{M}_{n_1}(D_1) \times \cdots \times \mathrm{M}_{n_r}(D_r).$$

Such a decomposition is unique up to permuting the integers $n_1, \dots, n_r$ and applying an isomorphism to the division rings $D_1, \dots, D_r$.

Proof. If $B \simeq \prod_i M_{n_i}(D_i)$, then each factor $M_{n_i}(D_i)$ is a simple F-algebra by Example 7.2.9, so by Corollary 7.3.8, B is semisimple.

So suppose B is semisimple. Then we can write B as a left B-module as the direct sum $B \simeq I_1^{\oplus n_1} \oplus \cdots \oplus I_r^{\oplus n_r}$ of simple B-modules $I_1, \dots, I_r$, grouped up to isomorphism. We have $\mathrm{End}_B(B) \simeq B^{\mathrm{op}}$ by 7.2.14. By Schur's lemma,

$$\mathrm{End}_B(B) \simeq \bigoplus_i \mathrm{End}_B\big(I_i^{\oplus n_i}\big);$$

by 7.2.16,

$$\operatorname{End}_B\left(I_i^{\oplus n_i}\right) \simeq \mathrm{M}_{n_i}(D_i)$$

where $D_i = \operatorname{End}_B(I_i)$ is a division ring. So

$$B \simeq \operatorname{End}_B(B)^{\mathrm{op}} \simeq \mathrm{M}_{n_1}(D_1^{\mathrm{op}}) \times \cdots \times \mathrm{M}_{n_r}(D_r^{\mathrm{op}}).$$

The statements about uniqueness are then clear. □

Remark 7.3.11. Main Theorem 7.3.10 as it is stated was originally proven by Wedderburn [Wed08], and so is sometimes called *Wedderburn's theorem*. However, this term may also apply to the theorem of Wedderburn that a finite division ring is a field; and Artin generalized Main Theorem 7.3.10 to rings where the ascending and descending chain condition holds for left ideals [Art26]. We follow the common convention by referring to Main Theorem 7.3.10 as the Wedderburn–Artin theorem.

Corollary 7.3.12. *Let B be a simple F-algebra. Then $B \simeq \mathrm{M}_n(D)$ for a unique $n \in \mathbb{Z}_{\geq 1}$ and a division algebra D unique up to isomorphism.*

Example 7.3.13. Let B be a division F-algebra. Then $V = B$ is a simple B-module, and in Corollary 7.3.12 we have $D = \operatorname{End}_B(B) = B^{\mathrm{op}}$, and the Wedderburn–Artin isomorphism is just $B \simeq \mathrm{M}_1((B^{\mathrm{op}})^{\mathrm{op}})$.

Corollary 7.3.14. *An F-algebra B is semisimple if and only if B is the direct product of simple F-algebras.*

Proof. Immediate from the Wedderburn–Artin theorem, as each factor $\mathrm{M}_{n_i}(D_i)$ is simple. □

7.4 Jacobson radical

We now consider an important criterion for establishing the semisimplicity of an F-algebra. Let B be a finite-dimensional F-algebra.

Definition 7.4.1. The **Jacobson radical** $\operatorname{rad} B$ of B is the intersection of all maximal left ideals of B.

We will in Corollary 7.4.6 see that this definition has left-right symmetry. Before doing so, we see right away the importance of the Jacobson radical in the following lemma.

Lemma 7.4.2. *B is semisimple if and only if* $\operatorname{rad} B = \{0\}$.

Proof. First, suppose B is semisimple. Then B as a left B-module is isomorphic to the direct sum of simple left ideals of B. Suppose $\operatorname{rad} B \neq \{0\}$; then $\operatorname{rad} B$ contains a minimal, hence simple, nonzero left ideal $I \subseteq B$. Then $B = I \oplus I'$ for some B-submodule I' and $B/I' \simeq I$ so I' is a maximal left ideal. Therefore $\operatorname{rad} B \subseteq I'$, but then $\operatorname{rad} B \cap I = \{0\}$, a contradiction.

Conversely, suppose $\operatorname{rad} B = \{0\}$. Suppose B is not semisimple. Let I_1 be a minimal left ideal of B. Since $I_1 \neq \{0\} = \operatorname{rad} B$, there exists a maximal left ideal J_1 not

containing I_1, so $I_1 \cap J_1 = \{0\}$ and $B = I_1 \oplus J_1$. Since B is not semisimple, $J_1 \neq \{0\}$, and there exists a minimal left ideal $I_2 \subsetneq J_1 \subseteq B$. Continuing in this fashion, we obtain a descending chain $J_1 \supsetneq J_2 \supsetneq \dots$, a contradiction. □

Corollary 7.4.3. *$B/\operatorname{rad} B$ is semisimple.*

Proof. Let $J = \operatorname{rad} B$. Under the natural map $B \to B/J$, the intersection of all maximal left ideals of $B/\operatorname{rad} B$ corresponds to the intersection of all maximal left ideals of B containing J; but $\operatorname{rad} B$ is the intersection thereof, so $\operatorname{rad}(B/J) = \{0\}$ and by Lemma 7.4.2, B/J is semisimple. □

We now characterize the Jacobson radical in several ways.

7.4.4. For a left B-module V, define its **annihilator** by

$$\operatorname{ann} V := \{\alpha \in B : \alpha V = 0\}.$$

Every annihilator $\operatorname{ann} V$ is a two-sided ideal of B: if $\alpha \in \operatorname{ann}(V)$ and $\beta \in B$, then $\alpha\beta V \subseteq \alpha V = \{0\}$ so $\alpha\beta \in \operatorname{ann}(V)$.

Lemma 7.4.5. *The Jacobson radical is equal to the intersection of the annihilators of all simple left B-modules: i.e., we have $\operatorname{rad} B = \bigcap_V \operatorname{ann} V$, the intersection taken over all simple left B-modules. Moreover, if $\alpha \in \operatorname{rad} B$, then $1 - \alpha \in B^\times$.*

Proof. We begin with the containment ($\supseteq$). Let $\alpha \in \bigcap_V \operatorname{ann} V$ and let I be a maximal left ideal. Then $V = B/I$ is a simple left B-module, so $\alpha \in \operatorname{ann}(B/I)$ whence $\alpha B \subseteq I$ and $\alpha \in I$.

The containment ($\subseteq$) follows with a bit more work. Let $\alpha \in \operatorname{rad} B$, and let V be a simple left B-module. Assume for purposes of contradiction that $x \in V$ has $\alpha x \neq 0$. Then as in Lemma 7.2.7, $V = B(\alpha x)$ so $x = \beta\alpha x$ for some $\beta \in B$ and $(1 - \beta\alpha)x = 0$. Let I be a maximal left ideal containing $1 - \beta\alpha$. Since $\alpha \in \operatorname{rad} B$, we have $\alpha \in I$, and thus $1 = (1 - \beta\alpha) + \beta\alpha \in I$, a contradiction. Thus $\alpha V = \{0\}$ and $\alpha \in \operatorname{ann} V$.

The final statement follows along similar lines as the previous paragraph (Exercise 7.10). □

Corollary 7.4.6. *The Jacobson radical $\operatorname{rad} B$ is a two-sided ideal of B.*

Proof. The statement follows by combining 7.4.4 and Lemma 7.4.5: $\operatorname{rad} B$ is the intersection of two-sided ideals and so is itself a two-sided ideal. □

Example 7.4.7. If B is commutative (and still a finite-dimensional F-algebra), then $\operatorname{rad} B = \sqrt{(0)}$ is the **nilradical** of B, the set of all nilpotent elements of B.

A two-sided ideal $J \subseteq B$ is **nilpotent** if $J^n = \{0\}$ for some $n \geq 1$, i.e., every product of n elements from J is zero. Every element of a nilpotent ideal is itself nilpotent.

Lemma 7.4.8. *$J = \operatorname{rad} B$ contains every nilpotent two-sided ideal, and J itself is nilpotent.*

Proof. If $I \subseteq B$ is a nilpotent two-sided ideal, then $I + J$ is a nilpotent two-sided ideal of B/J; but $\operatorname{rad}(B/J) = \{0\}$ by Corollary 7.4.3, so B/J is the direct product of simple algebras (Corollary 7.3.14) and therefore has no nonzero nilpotent two-sided ideals. Therefore $I \subseteq I + J \subseteq J$.

Now we prove that J is nilpotent. Consider the descending chain

$$B \supset J \supseteq J^2 \supseteq \dots.$$

There exists $n \in \mathbb{Z}_{\geq 1}$ such that $J^n = J^{2n}$. We claim that $J^n = \{0\}$. Assume for the purposes of contradiction that $I \subseteq J^n$ is a minimal left ideal such that $J^n I \neq \{0\}$. Let $\alpha \in I$ be such that $J^n\alpha \neq \{0\}$; by minimality $J^n\alpha = I$, so $\alpha = \eta\alpha$ for some $\eta \in J^n$, thus $(1 - \eta)\alpha = 0$. But $\eta \in J^n \subseteq J = \operatorname{rad} B$. By Lemma 7.4.5, $1 - \eta \in B^\times$ is a unit hence $\alpha = 0$, a contradiction. □

Example 7.4.9. Suppose B has a standard involution. Then by Lemma 7.4.8 and the fact that B has degree 2, we conclude that $\operatorname{rad} B \subseteq \{\epsilon \in B : \epsilon^2 = 0\}$. If $\operatorname{char} F \neq 2$ and we define $\operatorname{rad}(\mathrm{nrd})$ as in 4.3.9 for the quadratic form nrd, then $\operatorname{rad}(\mathrm{nrd}) = \operatorname{rad} B$ (Exercise 7.21).

Corollary 7.4.10. *The Jacobson radical* rad B *is the intersection of all maximal right ideals of* B.

Proof. Lemma 7.4.8 gives a left-right symmetric characterization of the Jacobson radical, so $\operatorname{rad} B = \operatorname{rad} B^{\mathrm{op}}$. There is a bijection between simple left B-modules and simple right B^{op}-modules, and the result follows. □

7.5 Central simple algebras

For more on central simple algebras (and in particular division algebras), see e.g. Saltman [Sal99] or Draxl [Dra83].

Recall (2.1.1) that the center of B is defined as

$$Z(B) := \{\alpha \in B : \alpha\beta = \beta\alpha \text{ for all } \alpha \in B\}.$$

Remark 7.5.1. An F-algebra B is a central $Z(B)$-algebra when $Z(B)$ is a field. (Under a more general definition of algebra, every algebra is an algebra over its center.)

Example 7.5.2. The center $Z(B)$ of a simple F-algebra is a field, since it is a simple commutative F-algebra. One reaches the same conclusion by applying Corollary 7.3.12 together with $Z(\mathrm{M}_n(D)) = Z(D)$ (Exercise 7.5).

The category of central simple algebras is closed under tensor product, as follows.

Proposition 7.5.3. *Let* A, B *be* F*-algebras and suppose that* B *is central.*

(a) *The center of* $A \otimes_F B$ *is the image of* $Z(A) \hookrightarrow A \otimes_F B$ *under* $z \mapsto z \otimes 1$.
(b) *Suppose that* A, B *are simple. Then* $A \otimes_F B$ *is simple.*

Proof. First, centrality in part (a). Suppose that $\gamma = \sum_i \alpha_i \otimes \beta_i \in Z(A \otimes B)$ (a finite sum). By rewriting the tensor, without loss of generality, we may suppose that α_i are linearly independent over F. Then by properties of tensor products, the elements $\beta_i \in B$ in the representation $\gamma = \sum_i \alpha_i \otimes \beta_i$ are unique. But then for all $\beta \in B$,

$$\sum_i (\alpha_i \otimes \beta\beta_i) = (1 \otimes \beta)\left(\sum_i \alpha_i \otimes \beta_i\right) = \left(\sum_i \alpha_i \otimes \beta_i\right)(1 \otimes \beta) = \sum_i (\alpha_i \otimes \beta_i\beta)$$

so $\beta\beta_i = \beta_i\beta$ for each i; thus $\beta_i = b_i \in Z(B) = F$. Hence

$$\gamma = \sum_i \alpha_i \otimes b_i = \sum_i \alpha_i b_i \otimes 1 = \left(\sum_i \alpha_i b_i\right) \otimes 1;$$

since $\alpha \otimes 1$ also commutes with γ for all $\alpha \in A$, we have $\sum_i \alpha_i b_i \in Z(A)$. Thus $\gamma \in Z(A) \otimes F = Z(A)$.

Next, simplicity in part (b). Let I be a nontrivial two-sided ideal in $A \otimes B$, and let $\gamma = \sum_{i=1}^m \alpha_i \otimes \beta_i \in I \smallsetminus \{0\}$. Without loss of generality, we may suppose $\beta_1 \neq 0$. Then $B\beta_1 B = B$ since B is simple; multiplying on the left and right by elements of $B \subseteq A \otimes B$, we may suppose further that $\beta_1 = 1$. Let $\gamma \in I \smallsetminus \{0\}$ be such an element that is minimal with respect to m; then in particular the elements β_i are linearly independent over F. Now for each $\beta \in B$,

$$(1 \otimes \beta)\gamma - \gamma(1 \otimes \beta) = \sum_{i=2}^m (\alpha_i \otimes (\beta\beta_i - \beta_i\beta)) \in I;$$

but by minimality of m, the right-hand side is zero, so $\beta\beta_i = \beta_i\beta$ for all i. Hence $\beta_i \in Z(B) = F$ for all i and as above $\gamma = \alpha \otimes 1$ for some $0 \neq \alpha \in A$. But then

$$I \supseteq (A \otimes 1)(\alpha \otimes 1)(A \otimes 1) = (A\alpha A) \otimes 1 = A \otimes 1$$

since A is simple, so $I \supseteq (A \otimes 1)(1 \otimes B) = A \otimes B$, and thus $I = A \otimes B$ and $A \otimes B$ is simple. □

Lemma 7.5.4. *If B is a finite-dimensional algebra over F, then B is a central simple F-algebra if and only if the map*

$$\begin{aligned} \phi\colon B \otimes_F B^{\mathrm{op}} &\xrightarrow{\sim} \mathrm{End}_F(B) \\ \textstyle\sum_i \alpha_i \otimes \beta_i &\mapsto (\mu \mapsto \textstyle\sum_i \alpha_i \mu \beta_i) \end{aligned}$$

is an isomorphism.

Proof. First, the implication (⇒). Just as in 7.2.14, ϕ is a nonzero F-algebra homomorphism. By Proposition 7.5.3, $B \otimes_F B^{\mathrm{op}}$ is simple, so ϕ is injective. Since $\dim_F(B \otimes_F B^{\mathrm{op}}) = \dim_F \mathrm{End}_F(B) = (\dim_F B)^2$, ϕ is an isomorphism.

Now the converse (⇐); suppose ϕ is an isomorphism. If I is an ideal of B then $\phi(I \otimes B^{\mathrm{op}}) \subseteq \mathrm{End}_F(B)$ is an ideal; but $\mathrm{End}_F(B)$ is simple over F, therefore I is trivial. And if $\alpha \in Z(B)$ then $\phi(\alpha \otimes 1) \in Z(\mathrm{End}_F(B)) = F$, so $\alpha \in F$. □

7.5.5. Among central simple algebras over a field, quaternion algebras have an especially nice presentation because the quadratic norm form can be put into a standard form (indeed, diagonalized in characteristic not 2). More generally, one may look at algebras with a similarly nice presentation, as follows.

Let F be a field, let $K \supset F$ be a cyclic extension of F of degree $n = [K : F]$, let $\sigma \in \mathrm{Gal}(K \,|\, F)$ be a generator, and let $b \in F^\times$. For example, if F contains a primitive nth root of unity $\zeta \in F^\times$, and $a \in F^\times \smallsetminus F^{\times n}$, then we may take $K = F(\sqrt[n]{a})$ and $\sigma(\sqrt[n]{a}) = \zeta \sqrt[n]{a}$. We then define the **cyclic algebra**

$$\left(\frac{K, \sigma, b}{F}\right) = K \oplus Kj \oplus \cdots \oplus Kj^{n-1}$$

to be the left K-vector space with basis $1, j, \dots, j^{n-1}$ and with multiplication $j^n = b$ and $j\alpha = \sigma(\alpha)j$ for $\alpha \in K$. The definition of a cyclic algebra generalizes that of 6.1.5, where there is only one choice for the generator σ. A cyclic algebra is a central simple algebra over F of dimension n^2, and indeed $(K, \sigma, b \mid K) \simeq \mathrm{M}_n(K)$. (See Exercise 7.12.) More generally, we may relax the condition that G be cyclic: there is an analogous construction for any finite Galois extension, yielding a central simple algebra called a *crossed product algebra* (and giving an interpretation to a second cohomology group): see Reiner [Rei2003, §29–30]. There are significant open problems relating cyclic algebras and crossed products to central simple algebras in general [ABGV2006].

It is a consequence of the main theorem of class field theory that if F is a global field then every (finite-dimensional) central simple algebra over F is isomorphic to a cyclic algebra.

Remark 7.5.6. The theory of central simple algebras and Brauer groups extends to one over commutative rings (or even schemes), and this becomes the theory of Azumaya algebras: see Saltman [Sal99, §2].

7.6 Quaternion algebras

Having set the stage, we are now ready to prove the following final characterizations of quaternion algebras.

Proposition 7.6.1. *Let B be an F-algebra. Then the following are equivalent:*

(i) *B is a quaternion algebra;*
(ii) *B is a central simple F-algebra with $\dim_F B = 4$;*
(iii) *B is a central semisimple F-algebra with $\dim_F B = 4$; and*
(iv) *$B \otimes_F F^{\mathrm{al}} \simeq \mathrm{M}_2(F^{\mathrm{al}})$, where F^{al} is an algebraic closure of F.*

Proof. First, (i) ⇒ (ii): if B is a quaternion algebra, then B is central simple (7.2.11).

The equivalence (ii) ⇔ (iii) follows from the Wedderburn–Artin theorem:

$$1 = \dim Z(B) = \sum_{i=1}^{r} \dim_F Z(D_i) \geq r$$

so $r = 1$.

Next we prove (ii) $\Rightarrow$ (iv). If B is central simple, then $B \otimes_F F^{\mathrm{al}}$ is a central simple F^{al}-algebra by Proposition 7.5.3. But by Exercise 2.9, the only division F^{al}-algebra is F^{al}, so by the Wedderburn–Artin theorem, $B \otimes_F F^{\mathrm{al}} \simeq \mathrm{M}_n(F^{\mathrm{al}})$; by dimensions, $n = 2$.

It remains to prove (iv) $\Rightarrow$ (i). So suppose $B \otimes_F F^{\mathrm{al}} \simeq \mathrm{M}_2(F^{\mathrm{al}})$. Then B is simple by Example 7.2.10 and $\dim_F B = 4$. By the Wedderburn–Artin theorem (Corollary 7.3.12), we have $B \simeq \mathrm{M}_n(D)$ with $n \in \mathbb{Z}_{\geq 1}$ and D a division ring. Since $4 = \dim_F B = n^2 \dim_F D$, either $n = 2$ and $B \simeq \mathrm{M}_2(F)$, or $n = 1$ and B is a division ring.

In this latter case, the result will follow from Theorem 3.5.1 (and Theorem 6.2.8 for the case $\operatorname{char} F = 2$) if we show that B has degree 2. But for any $\alpha \in B$ we have that $\alpha \in B \otimes_F F^{\mathrm{al}} \simeq \mathrm{M}_2(F^{\mathrm{al}})$ satisfies its characteristic polynomial of degree 2, so that $1, \alpha, \alpha^2$ are linearly dependent over F^{al} and hence linearly dependent over F, by linear algebra. □

Inspired by the proof of this result, we reconsider and reprove our splitting criterion for quaternion algebras.

Proposition 7.6.2. *Let B be a quaternion algebra over F. Then the following are equivalent:*

(i) *$B \simeq \mathrm{M}_2(F)$;*
(ii) *B is not a division ring;*
(iii) *There exists $0 \neq \epsilon \in B$ such that $\epsilon^2 = 0$;*
(iv) *B has a nontrivial left ideal $I \subsetneq B$;*

Proof. The equivalence (i) $\Leftrightarrow$ (ii) follows from the Wedderburn–Artin theorem (also proved in Main Theorem 5.4.4 and Theorem 6.4.11). The implications (i) $\Rightarrow$ (iii) $\Rightarrow$ (ii) and (i) $\Rightarrow$ (iv) $\Rightarrow$ (ii) are clear. □

7.6.3. We showed in Lemma 7.2.17 that a simple algebra B has a unique simple left B-module I up to isomorphism, obtained as a minimal nonzero left ideal. If B is a quaternion algebra, this simple module I can be readily identified using the above proposition. If B is a division ring, then necessarily $I = B$. Otherwise, $B \simeq \mathrm{M}_2(F)$, and then $I \simeq F^2$, and the map $B \to \operatorname{End}_F(I)$ given by left matrix multiplication is an isomorphism.

7.7 The Skolem–Noether theorem

We conclude this chapter with a fundamental result that characterizes the automorphisms of a simple algebra—and much more.

Main Theorem 7.7.1. (Skolem–Noether). *Let A, B be simple F-algebras and suppose that B is central. Suppose that $f, g \colon A \to B$ are homomorphisms. Then there exists $\beta \in B^\times$ such that $f(\alpha) = \beta^{-1} g(\alpha) \beta$ for all $\alpha \in A$.*

Proof. By Corollary 7.3.12, we have $B \simeq \operatorname{End}_D(V) \simeq \mathrm{M}_n(D^{\mathrm{op}})$ where V is a simple B-module and $D = \operatorname{End}_B(V)$ is a central F-algebra. Now the maps f, g give V the structure of an A-module in two ways. The A-module structure commutes with the D-module structure since $B \simeq \operatorname{End}_D(V)$. So V has two $A \otimes_F D$-module structures via f and g.

By Proposition 7.5.3, since D is central over F, we have that $A \otimes_F D$ is a simple F-algebra. By Corollary 7.3.9 and a dimension count, the two $A \otimes_F D$-module structures on V are isomorphic. Thus, there exists an isomorphism $\beta \colon V \to V$ of $A \otimes_F D$-modules; i.e. $\beta(f(\alpha)x) = g(\alpha)\beta(x)$ for all $\alpha \in A$ and $x \in V$, and $\beta(\delta x) = \delta\beta(x)$ for all $\delta \in D$ and $x \in V$. We have $\beta \in \operatorname{End}_D(V) \simeq B$ and so we can write $\beta f(\alpha)\beta^{-1} = g(\alpha)$ for all $\alpha \in A$, as claimed. □

The following corollaries are immediate consequences (special cases) of the Skolem–Noether theorem.

Corollary 7.7.2. *If A_1, A_2 are simple F-subalgebras of a central simple F-algebra B and $\phi \colon A_1 \xrightarrow{\sim} A_2$ is an isomorphism of F-algebras, then ϕ is induced by an inner automorphism of B.*

Proof. Let $\iota_i \colon A_i \hookrightarrow B$ be the natural inclusions, and apply Main Theorem 7.7.1 to $f := \iota_1$ and $g := \iota_2 \circ \phi$: we conclude there exists $\beta \in B^\times$ such that $\iota_1(\alpha) = \alpha = \beta^{-1}\iota_2(\phi(\alpha))\beta$ or equivalently $\phi(\alpha) = \beta\alpha\beta^{-1}$ for all $\alpha \in A_1$, as desired. □

Corollary 7.7.3. *If B is a central simple F-algebra and $\alpha_1, \alpha_2 \in B^\times$, then α_1, α_2 have the same irreducible minimal polynomial over F if and only if there exists $\beta \in B^\times$ such that $\alpha_2 = \beta^{-1}\alpha_1\beta$.*

Proof. The implication ($\Leftarrow$) is immediate. Conversely ($\Rightarrow$), let $A_i := F[\alpha_i] \simeq F[x]/(f_i(x))$ where $f_i(x) \in F[x]$ are minimal polynomials over F. Since these polynomials are irreducible, A_i is a field hence simple, so Corollary 7.7.2 gives the result. □

Corollary 7.7.4. *The group of F-algebra automorphisms of a central simple algebra B is* $\operatorname{Aut}(B) \simeq B^\times/F^\times$.

Proof. Taking $A = B$ in Main Theorem 7.7.1, we conclude that every automorphism of B as an F-algebra is inner, and an inner automorphism is trivial if and only if it is conjugation by an element of the center $F^\times$. □

Example 7.7.5. By Corollary 7.7.4, we have a canonical isomorphism of groups

$$\operatorname{Aut}(\mathrm{M}_n(F)) \cong \mathrm{GL}_n(F)/F^\times = \mathrm{PGL}_n(F).$$

As a final application, we extend the splitting criterion of Main Theorem 5.4.4(i) $\Leftrightarrow$ (vi) to detect *isomorphism classes* of quaternion algebras.

Corollary 7.7.6. *Let $K \supseteq F$ be a separable quadratic F-algebra, and let $b, b' \in F^\times$. Then*

$$\left(\frac{K, b}{F}\right) \simeq \left(\frac{K, b'}{F}\right) \quad \Leftrightarrow \quad b/b' \in \operatorname{Nm}_{K|F}(K^\times).$$

Taking $b' = 1$, we recover the previous splitting criteria.

Proof. For the implication ($\Leftarrow$), if $b'/b = \mathrm{Nm}_{K|F}(\alpha)$ with $\alpha \in K^\times$, then an isomorphism is furnished as left K-vector spaces by sending $j \mapsto \alpha j$.

For the implication ($\Rightarrow$), let $\phi\colon (K, b \mid F) \xrightarrow{\sim} B' := (K, b' \mid F)$ be an isomorphism of F-algebras. If $K \simeq F \times F$ is not a field, then $\mathrm{Nm}_{K|F}(K^\times) = F^\times$ and the result holds. So suppose K is a field. Then $\phi(K) \subseteq B'$ isomorphic to K as an F-algebra, but need not be the designated one in B'; however, by the Skolem–Noether theorem, we may postcompose ϕ with an automorphism that sends $\phi(K)$ to the designated one, i.e., we may suppose that ϕ is a K-linear map (taking the algebras as left K-vector spaces). Let $\phi(j) = \alpha + \beta j'$ with $\alpha, \beta \in K$. Then

$$\{0\} = \mathrm{trd}(Kj) = \mathrm{trd}(\phi(Kj)) = \mathrm{trd}(K\phi(j)) = \mathrm{trd}(K\alpha)$$

and thus $\alpha = 0$ since K is separable. Consequently,

$$-b = \mathrm{nrd}(j) = \mathrm{nrd}(\phi(j)) = -\mathrm{Nm}_{K|F}(\beta)b'$$

and so $b/b' = \mathrm{Nm}_{K|F}(\beta)$ as desired. $\square$

In the remainder of this section, we prove an important consequence of the Skolem–Noether theorem that compares centralizers of subalgebras to dimensions.

Definition 7.7.7. Let A be an F-subalgebra of B. Let

$$C_B(A) := \{\beta \in B : \alpha\beta = \beta\alpha \text{ for all } \alpha \in A\}$$

be the **centralizer** of A in B.

The centralizer $C_B(A)$ is an F-subalgebra of B.

Proposition 7.7.8. *Let B be a central simple F-algebra and let $A \subsetneq B$ a simple F-subalgebra. Then the following statements hold:*

(a) $C_B(A)$ *is a simple F-algebra.*
(b) $\dim_F B = \dim_F A \cdot \dim_F C_B(A)$.
(c) $C_B(C_B(A)) = A$.

Part (c) of this proposition is called the **double centralizer property**.

Proof. First, part (a). We interpret the centralizer as arising from certain kinds of endomorphisms. We have that B is a left $A \otimes B^{\mathrm{op}}$ module by the action $(\alpha \otimes \beta) \cdot \mu = \alpha\mu\beta$ for $\alpha \otimes \beta \in A \otimes B^{\mathrm{op}}$ and $\mu \in B$. We claim that

$$C_B(A) = \mathrm{End}_{A \otimes B^{\mathrm{op}}}(B). \tag{7.7.9}$$

Any $\phi \in \mathrm{End}_{A\otimes B^{\mathrm{op}}}(B)$ is left multiplication by an element of B: if $\gamma = \phi(1)$, then $\phi(\mu) = \phi(1)\mu = \gamma\mu$ by $1 \otimes B^{\mathrm{op}}$-linearity. Now the equality

$$\gamma\alpha = \phi(\alpha) = \alpha\phi(1) = \alpha\gamma$$

shows that multiplication by γ is $A \otimes 1$-linear if and only if $\gamma \in C_B(A)$, proving (7.7.9).

By Proposition 7.5.3, the algebra $A \otimes B^{\mathrm{op}}$ is simple. By the Wedderburn–Artin theorem, $A \otimes B^{\mathrm{op}} \simeq \mathrm{M}_n(D)$ for some $n \geq 1$ and division F-algebra D. Since $\mathrm{M}_n(D)$ is simple, its unique simple left D-module is $V = D^n$, and $\mathrm{End}_{\mathrm{M}_n(D)}(V) \simeq D^{\mathrm{op}}$. In particular, $B \simeq V^r$ for some $r \geq 1$ as an $A \otimes B^{\mathrm{op}}$-module. So

$$C_B(A) = \mathrm{End}_{A\otimes B^{\mathrm{op}}}(B) \simeq \mathrm{End}_{\mathrm{M}_n(D)}(V^r) \simeq \mathrm{M}_r(\mathrm{End}_{\mathrm{M}_n(D)}(V)) \simeq \mathrm{M}_r(D^{\mathrm{op}}).$$

Thus $C_B(A)$ is simple.

For part (b),

$$\dim_F C_B(A) = \dim_F \mathrm{M}_r(D^{\mathrm{op}}) = r^2 \dim_F D$$

and

$$\dim_F(A \otimes B^{\mathrm{op}}) = \dim_F A \cdot \dim_F B = n^2 \dim_F D$$

and finally

$$\dim_F B = \dim_F V^r = r \dim_F D^n = rn \dim_F D;$$

putting these together gives $\dim_F A \cdot \dim_F C_B(A) = rn \dim_F D = \dim_F B$.

Finally, part (c) follows from (a) and (b):

$$\dim_F B = \dim_F C_B(A) \cdot \dim_F C_B(C_B(A)) = \dim_F A \cdot \dim_F C_B(A)$$

so $\dim_F A = \dim_F C_B(C_B(A))$ and $A \subseteq C_B(C_B(A))$, therefore equality holds. □

Example 7.7.10. We always have the two extremes $A = F$ and $A = B$, with $C_B(F) = B$ and $C_B(B) = F$, accordingly.

We note the following structurally crucial corollary of Proposition 7.7.8.

Corollary 7.7.11. *Let B be a central simple F-algebra and let K be a maximal subfield. Then $[B : F] = [K : F]^2$.*

Proof. If K is maximal, then $C_B(K) = K$ and $[B : F] = [K : F]^2$. □

Corollary 7.7.11 generalizes the comparatively easier statement for quaternion algebras: the maximal subfields of a quaternion algebra are quadratic. Returning now to quaternion algebras, we conclude with a nice package of consequences of the above results concerning *embeddings* of quadratic fields into quaternion algebras.

7.7.12. Let B be a quaternion algebra over F and let $K \subseteq B$ be a quadratic separable F-subalgebra. Then the set of all embeddings of K in B is naturally identified with the set $K^\times\backslash B^\times$, as follows.

By the Skolem–Noether theorem (Corollary 7.1.5, and Exercise 7.11 for the case $K \simeq F \times F$), if $\phi\colon K \hookrightarrow B$ is another embedding, then there exists $\beta \in B^\times$ such that $\phi(\alpha) = \beta^{-1}\alpha\beta$ for all $\alpha \in K$, and conversely. Such a conjugate embedding is the identity if and only if β centralizes K. By Corollary 4.4.5, and Corollary 6.4.2 for characteristic 2, the centralizer of $K^\times$ in $B^\times$ is $K^\times$. Therefore, the set of embeddings of K in B is naturally identified with the set $K^\times\backslash B^\times$, with $K^\times$ acting on the *left*.

7.8 Reduced trace and norm, universality

In this last section, we consider notions of reduced trace and reduced norm in the context of semisimple algebras. Thr

7.8.1. Let B be a (finite-dimensional) central simple algebra over F, and let F^{sep} denote a separable closure of F. By Exercise 7.24, we have an F-algebra homomorphism

$$\phi\colon B \otimes_F F^{\mathrm{sep}} \simeq \mathrm{M}_n(F^{\mathrm{sep}})$$

for some $n \geq 1$. By the Skolem–Noether theorem (Main Theorem 7.7.1), for any other isomorphism $\phi'\colon B \otimes_F F^{\mathrm{sep}} \simeq \mathrm{M}_n(F^{\mathrm{sep}})$, there exists $M \in \mathrm{GL}_n(F^{\mathrm{sep}})$ such that $\phi'(\alpha) = M\phi(\alpha)M^{-1}$, so the characteristic polynomial of an element of $B \otimes_F F^{\mathrm{sep}}$ is independent of the choice of ι. In particular, from the canonical embedding $\iota\colon B \hookrightarrow B \otimes_F F^{\mathrm{sep}}$ by $\alpha \mapsto \alpha \otimes 1$, we define the **reduced characteristic polynomial** of $\alpha \in B$ to be the characteristic polynomial of $(\phi\iota)(\alpha)$ as an element of $F^{\mathrm{sep}}[T]$ and similarly the **reduced trace** and **reduced norm** of α to be the trace and determinant of $(\phi\iota)(\alpha)$ as elements of F^{sep}.

In fact, the reduced characteristic polynomial descends to F, as follows. The absolute Galois group $\mathrm{Gal}_F := \mathrm{Gal}(F^{\mathrm{sep}} \mid F)$ acts on $B \otimes_F F^{\mathrm{sep}} \simeq \mathrm{M}_n(F^{\mathrm{sep}})$ by

$$\sigma(\alpha \otimes a) = \alpha \otimes \sigma(a)$$

for $\sigma \in \mathrm{Gal}_F$, $\alpha \in B$, and $a \in F^{\mathrm{sep}}$. Let $\sigma \in \mathrm{Gal}_F$. Since $\sigma(\alpha \otimes 1) = \alpha \otimes \sigma(1) = \alpha \otimes 1$, the reduced characteristic polynomials of $\iota(\alpha)$ and $\sigma(\iota(\alpha))$ are the same. By comparison (see e.g. Reiner [Rei2003, Theorem 9.3]), if

$$f(\alpha; T) = \det(T - \iota(\alpha)) = T^n + a_{n-1}T^{n-1} + \cdots + a_0$$

is the reduced characteristic polynomial of $\iota(\alpha)$, then the reduced characteristic polynomial of $(\sigma(\iota))(\alpha)$ is

$$\sigma(f)(\alpha; T) = \det(T - \sigma(\iota(\alpha))) = T^n + \sigma(a_{n-1})T^{n-1} + \cdots + \sigma(a_0).$$

And then since $f(\alpha; T) = \sigma(f)(\alpha; T)$ for all $\sigma \in \mathrm{Gal}_F$, by the fundamental theorem of Galois theory, $f(\alpha; T) \in F[T]$. Therefore, the reduced norm and reduced trace also belong to F.

Alternatively, we may argue as follows. The characteristic polynomial of left multiplication by α on B is the same as left multiplication by $(\phi\iota)(\alpha)$ on $\mathrm{M}_n(F^{\mathrm{sep}})$ (by extension of basis), and the latter is the nth power of the reduced characteristic polynomial by Exercise 3.13. Finally, if $f(T) \in F^{\mathrm{sep}}[T]$ has $f(T)^n \in F[T]$ then in fact $f(T) \in F[T]$: see Exercise 7.25.

These definitions extend to a general semisimple algebra over F, but to do so it is convenient to give an alternate approach that avoids going to the separable closure and works in even more generality using universal elements; for more, see Garibaldi [Gar2004].

Let B be a (finite-dimensional) F-algebra with $n := \dim_F B$, and choose a basis $e_1, \dots, e_n$ for B. Let $F(x_1, \dots, x_n)$ be a pure transcendental field extension of transcendence degree n, and let $\xi := x_1e_1 + \cdots + x_ne_n \in B \otimes_F F(x_1, \dots, x_n)$; we call ξ the **universal element** of B in the given basis.

Definition 7.8.2. The **universal minimal polynomial** of B (in the basis $e_1, \dots, e_n$) is the minimal polynomial $m_B(\xi;T)$ of ξ over $F(x_1, \dots, x_n)$.

For $\alpha = a_1 e_1 + \dots + a_n e_n \in B$ (with $a_i \in F$), the polynomial obtained from $m_B(\xi;T)$ by the substitution $x_i \leftarrow a_i$ is called the **specialization** of $m_B(\xi;T)$ at α.

The following example will hopefully illustrate the role of this notion.

Example 7.8.3. For $\operatorname{char} F \neq 2$ and $B = \left(\dfrac{a,b}{F}\right)$, in the basis $1, i, j, ij$ we have $\xi = t + xi + yj + zij$ (substituting t, x, y, z for $x_1, \dots, x_4$). We claim that the universal minimal polynomial is

$$m_B(\xi;T) = T^2 - 2tT + (t^2 - ax^2 - by^2 + abz^2).$$

Indeed, we verify that ξ satisfies $m_B(\xi;\xi) = 0$ by considering $\xi \in B \otimes_F F(t,x,y,z) = \left(\dfrac{a,b}{F(t,x,y,z)}\right)$ and computing that $\operatorname{trd}(\xi) = 2t$ and $\operatorname{nrd}(\xi) = t^2 - ax^2 - by^2 + abz^2$; and this polynomial is minimal because $\xi \notin F(t,x,y,z)$ does not satisfy a polynomial of degree 1 over $F(t,x,y,z)$.

7.8.4. If $B \simeq B_1 \times \dots \times B_r$, then in a basis for B obtained from the union of bases for the factors B_i with universal elements ξ_i, we have

$$m_B(\xi;T) = m_{B_1}(\xi_1;T) \cdots m_{B_r}(\xi_r;T).$$

In the proofs that follow, we abbreviate by using multi-index notation, e.g. writing $F[x] := F[x_1, \dots, x_n]$.

Lemma 7.8.5. *We have $m_B(\xi;T) \in F[x_1, \dots, x_n][T]$, i.e., the universal minimal polynomial has coefficients in $F[x_1, \dots, x_n]$.*

Proof. For part (a), we consider the map given by left multiplication by ξ on $B \otimes_F F(x)$. In the basis $e_1, \dots, e_n$, almost by construction we find that the matrix of this map has coefficients in $F[x]$ (it is the matrix of linear forms obtained from left multiplication by e_i). We conclude that ξ satisfies the characteristic polynomial of this matrix, which is a monic polynomial with coefficients in $F[x]$. Since $m_B(\xi;T)$ divides this polynomial (over $F(x)$) by minimality, by Gauss's lemma we conclude that $m_B(\xi;T) \in F[x][T]$. ◻

Proposition 7.8.6. *For all $\alpha \in B$, the specialization of $m_B(\xi;T)$ at α is independent of the choice of basis $e_1, \dots, e_n$ and is satisfied by the element α. Moreover, if $\phi \in \operatorname{Aut}(B)$ and $\alpha \in B$, then α and $\phi(\alpha)$ have the same specialized polynomials.*

In view of Proposition 7.8.6, we write $m_B(\alpha;T)$ for the specialization of $m_B(\xi;T)$ at $\alpha \in B$; from it, we conclude that $m_B(\alpha;\alpha) = 0$.

Proof. Since $m_B(\xi;\xi) = 0$, by specialization we obtain $m_B(\alpha;\alpha) = 0$. For the independence of basis, let $e_1', \dots, e_n'$ be another F-basis and ξ' the corresponding universal element. Writing e_i in the basis e_i' allows us to write $\xi = \sum_{i=1}^n \ell_i(x) e_i'$ where

$\ell_i(x) \in F[x]$ are linear forms; moreover, writing $\alpha = \sum_i a_i' e_i'$ we have $\ell_i(a) = a_i'$. The map $x_i \mapsto \ell_i(x)$ extends to an F-algebra automorphism ϕ of $F[x]$ (repeat the construction with the inverse, and compose) with $\phi(c)(a) = c(a')$ for all $c(x) \in F[x]$. We let ϕ act on polynomials over $F[x]$ by acting on the coefficients; by uniqueness of minimal polynomials, we have $\phi(m_B(\xi;T)) = m_B(\xi';T)$. Therefore, looking at each coefficient, specializing $\phi(m_B(\xi;T))$ at a is the same as specializing $m_B(\xi';T)$ at a', as claimed.

The second sentence follows by the same argument, as from $\phi \in \operatorname{Aut}(B)$ we have a new basis $e_i' := \phi(e_i)$ so the specializations again agree. (This argument replaces the use of the Skolem–Noether theorem in the special case where B is a central simple algebra.) □

Lemma 7.8.7. *For any field extension $K \supseteq F$, we have $m_{B\otimes_F K}(\xi;T) = m_K(\xi;T)$.*

Proof. First, because an F-basis for B is a K-basis for $B \otimes_F K$, the element ξ (as the universal element of B), also serves as a universal element of $B \otimes_F K$. Since $K(x_1, \dots, x_n) \subseteq F(x_1, \dots, x_n)$, by minimality we have $m_{B_K}(\xi;T) \mid m_B(\xi;T)$. Conversely, let $F(x)[\xi] \subseteq B \otimes_F F(x)$ be the subalgebra generated by ξ over $F(x)$; then $F(x)[\xi] \simeq F(x)[T]/(m_B(\xi;T))$. Tensoring with K gives $K(x)[\xi] \simeq K(x)[T]/(m_B(\xi;T))$ as the subalgebra of $(B \otimes_F K) \otimes_K K(x)$ generated by ξ. Thus $m_B(\xi;T) \mid m_{B_K}(\xi;T)$, so equality holds. □

We conclude by relating this construction to more familiar polynomials.

Lemma 7.8.8. *Let $\alpha \in B$. Then the following statements hold.*

(a) *If $B = K \supseteq F$ is a separable field extension, then $m_K(\alpha;T)$ is the characteristic polynomial of left multiplication by α.*
(b) *If B is a central simple F-algebra, then $m_B(\alpha;T)$ is the reduced characteristic polynomial of α.*

Proof. For (a), we recall (as in the proof of Lemma 7.8.5 that ξ satisfies the characteristic polynomial of left multiplication on K, a polynomial of degree $n = [K : F]$; on the other hand, choosing a primitive element α, we see the specialization $m_K(\alpha;T)$ is satisfied by α so has degree at least n, so equality holds and $m_K(\alpha;T)$ is the characteristic polynomial. Therefore the universal minimal polynomial is the characteristic polynomial, and hence the same is true under every specialization.

For (b), it suffices to prove this when $F = F^{\mathrm{al}}$ is algebraically closed, in which case $B \simeq \mathrm{M}_n(F)$; by Proposition 7.8.6, we may assume $B = \mathrm{M}_n(F)$. By 7.8.1, we want to show that $m_B(\alpha;T)$ for $\alpha \in \mathrm{M}_n(F)$ is the usual characteristic polynomial. But the universal element (in a basis of matrix units, or any basis) satisfies its characteristic polynomial of degree n, and a nilpotent matrix (with 1s just above the diagonal) has minimal polynomial T^n, so we conclude as in the previous paragraph. □

In light of the above, we may make the following definition.

Definition 7.8.9. Let B be a semisimple F-algebra. For $\alpha \in B$, the **reduced characteristic polynomial** $f(\alpha;T) = T^n - c_1T^{n-1} + \cdots + (-1)^n c_n \in F[T]$ is the specialization of the universal minimal polynomial $m_B(\xi;T)$, and the **reduced trace** and **reduced norm** are the coefficients c_1, c_n, respectively.

Example 7.8.10. For a semisimple algebra $B \simeq B_1 \times \cdots \times B_r$, with each B_i central simple, we find that the reduced characteristic polynomial is just the product of the reduced characteristic polynomials on each simple direct factor B_i; this is well-defined (again) by the uniqueness statement in the Wedderburn–Artin theorem (Main Theorem 7.3.10).

Proposition 7.8.11. *Let B be semisimple. Then the reduced trace* $\operatorname{trd}\colon B \to F$ *is F-linear and satisfies* $\operatorname{trd}(\alpha\beta) = \operatorname{trd}(\beta\alpha)$ *for all $\alpha, \beta \in B$; and the reduced norm* $\operatorname{nrd}\colon B \to F$ *is multiplicative, satisfying* $\operatorname{nrd}(\alpha\beta) = \operatorname{nrd}(\alpha)\operatorname{nrd}(\beta)$.

Proof. Consider (again) $V := F(x)[\xi] \subseteq B \otimes_F F(x)$ the subalgebra generated over $F(x)$ by ξ; then ξ acts on $V \simeq F(x)[T]/(m_B(\xi;T))$ by left multiplication with characteristic polynomial $m_B(\xi;T)$. It follows that the reduced trace and reduced norm are the usual trace and determinant in this representation, so the announced properties follow on specialization. □

Remark 7.8.12. It is also possible to define the reduced characteristic polynomial on a semisimple algebra B by writing $B \simeq B_1 \times \cdots \times B_r$ as a product of simple algebras; for details, see Reiner [Rei2003, §9].

7.9 Separable algebras

For a (finite-dimensional) F-algebra, the notions of simple and semisimple are sensitive to the base field F in the sense that these properties need not hold after extending the base field. Indeed, let $K \supseteq F$ be a finite extension of fields, so K is a simple F-algebra. Then $K \otimes_F F^{\mathrm{al}}$ is simple only when $K = F$ and is semisimple if and only if $K \otimes_F F^{\mathrm{al}} \simeq F^{\mathrm{al}} \times \cdots \times F^{\mathrm{al}}$, i.e., K is separable over F.

It is important to have a notion which is stable under base change, as follows. For further reference, see Drozd–Kirichenko [DK94, §6], Curtis–Reiner [CR81, §7], Reiner [Rei2003, §7c], or Pierce [Pie82, Chapter 10].

Definition 7.9.1. Let B be a finite-dimensional F-algebra. We say that B is a **separable** F-algebra if B is semisimple and $Z(B)$ is a separable F-algebra.

In particular, a separable algebra over a field F with $\operatorname{char} F = 0$ is just a semisimple algebra.

7.9.2. For a semisimple algebra $B \simeq \mathrm{M}_{n_1}(D_1) \times \cdots \times \mathrm{M}_{n_r}(D_r)$, by Example 7.5.2 we have $Z(B) \simeq Z(D_1) \times \cdots \times Z(D_r)$, and B is separable if and only if $Z(D_i)$ is separable for each $i = 1, \ldots, r$.

Lemma 7.9.3. *A finite-dimensional simple F-algebra is a separable algebra over its center K.*

Proof. The center of B is a field $K = Z(B)$ and as a K-algebra, the center $Z(B) = K$ is certainly separable over K. (Or use Proposition 7.5.3 and Theorem 7.9.4(iii) below.) □

The notion of separability in this context is quite robust.

Theorem 7.9.4. *Let B be a finite-dimensional F-algebra. Then the following are equivalent:*

(i) *B is separable;*
(ii) *There exists a finite separable field extension K of F such that $B \otimes_F K \simeq M_{n_1}(K) \times \cdots \times M_{n_r}(K)$ for integers $n_1, \dots, n_r \geq 1$;*
(iii) *For every extension $K \supseteq F$ of fields, the K-algebra $B \otimes_F K$ is semisimple;*
(iv) *B is semisimple and the bilinear form*

$$\begin{aligned} B \times B &\to F \\ (\alpha, \beta) &\mapsto \operatorname{trd}(\alpha\beta) \end{aligned}$$

is nondegenerate.

Moreover, if $\operatorname{char} F = 0$*, then these are further equivalent to:*

(v) *The bilinear form $(\alpha, \beta) \mapsto \operatorname{Tr}_{B|F}(\alpha\beta)$ is nondegenerate.*

A separable F-algebra is sometimes called **absolutely semisimple**, in view of Theorem 7.9.4(iii).

Proof. First we prove (i) ⇒ (ii). Let B_i be a simple component of B; then $Z(B_i)$ is separable over F. Let $K_i \supseteq F$ be a separable field extension containing $Z(B_i)$ that splits B_i, so $B_i \otimes_{Z(B_i)} K_i \simeq \mathrm{M}_{n_i}(K_i)$. Let K be the compositum of the fields K_i. Then K is separable, and

$$B_i \otimes_F K \simeq \mathrm{M}_{n_i}(Z(B_i) \otimes_F K) \simeq \mathrm{M}_{n_i}(K) \times \cdots \times \mathrm{M}_{n_i}(K)$$

the number of copies equal to $[Z(B_i) : F]$.

Next we prove (ii) ⇒ (iii). Suppose $B \otimes_F K \simeq \prod_i \mathrm{M}_{n_i}(K)$ and let $L \supseteq F$ be an extension of fields. Let $M = KL$. On the one hand, $B \otimes_F M \simeq (B \otimes_F K) \otimes_K M \simeq \prod_i \mathrm{M}_{n_i}(M)$, so $\operatorname{rad} B \otimes_F M = \{0\}$; on the other hand, $B \otimes_F M \simeq (B \otimes_F L) \otimes_L M$ and $\operatorname{rad}(B \otimes_F L) \subseteq \operatorname{rad}(B \otimes_F L) \otimes_L M = \{0\}$, so $B \otimes_F L$ is semisimple.

For the implication (iii) ⇒ (i), suppose B is not separable, and we show that there exists $K \supseteq F$ such that $B \otimes_F K$ is not semisimple. If B is not semisimple over F, we can just take $F = K$. Otherwise, $Z(B)$ is not separable as an F-algebra, and there is a component of $Z(B)$ which is an inseparable field extension K. Then $B \otimes_F K$ contains a nonzero nilpotent element in its center and this element generates a nonzero nilpotent ideal, so $\operatorname{rad}(B \otimes_F K) \neq \{0\}$ and $B \otimes_F K$ is not semisimple.

The implication (iii) ⇒ (iv) holds for the following reason. We have $B \otimes_F F^{\mathrm{al}} \simeq \mathrm{M}_{n_1}(F^{\mathrm{al}}) \times \cdots \times \mathrm{M}_{n_r}(F^{\mathrm{al}})$, and the reduced trace pairing on each matrix ring factor is nondegenerate so the whole pairing is nondegenerate. By linear algebra we conclude that the bilinear form on B is nondegenerate.

The implication (iv) $\Rightarrow$ (i) holds with char F arbitrary: if $\epsilon \in \operatorname{rad} B$ then $\alpha\epsilon \in \operatorname{rad} B$ is nilpotent and $\operatorname{trd}(\alpha\epsilon) = 0$ for all $\alpha \in B$, and by nondegeneracy $\epsilon = 0$.

The final equivalence (iv) $\Leftrightarrow$ (v) follows when char $F = 0$ since the algebra trace pairing on each simple factor is a scalar multiple of the reduced trace pairing. □

Exercises

Throughout the exercises, let F be a field.

▹ 1. Prove that a quaternion algebra $B = \left(\dfrac{a,b}{F}\right)$ with char $F \neq 2$ is simple by a direct calculation, as follows.

   (a) Let I be a nontrivial two-sided ideal, and let $\epsilon = t + xi + yj + zij \in I$. By considering $i\epsilon - \epsilon i$, show that $t + xi \in I$.

   (b) Arguing symmetrically and taking a linear combination, show that $t \in I$, and conclude that $t = 0$, whence $x = y = z = 0$.

   Modify this argument to show that an algebra $B = \left[\dfrac{a,b}{F}\right)$ is simple when char $F = 2$. [We proved these statements without separating into cases in 7.2.11.]

2. Let B be a quaternion algebra over F, and let $K \subseteq B$ be an F-subalgebra that is *commutative*. Show that $\dim_F K \leq 2$.

3. Let B be a quaternion algebra. Exhibit an explicit isomorphism

$$B \otimes_F B \xrightarrow{\sim} \mathrm{M}_4(F).$$

   [Hint: see Exercise 2.11.]

▹ 4. Let $B = \left\{\begin{pmatrix} a & b \\ 0 & c \end{pmatrix} : a, b, c \in F\right\} \subseteq \mathrm{M}_2(F)$, and $V = F^2$ be the left B-module of column vectors. Show that V is indecomposable, but not simple, as a left B-module (cf. Example 7.2.5).

▹ 5. This exercise proves basic but important facts about two-sided ideals in matrix algebras using matrix units.

   (a) Let D be a division F-algebra. Prove that $\mathrm{M}_n(D)$ is a simple F-algebra with center $Z(D)$ for all $n \geq 1$. *[Hint: Let E_{ij} be the matrix with 1 in the ijth entry and zeros in all other entries. Show that $E_{ki} M E_{j\ell} = m_{ij} E_{k\ell}$ where m_{ij} is the ijth entry of M.]*

   (b) More generally, let R be a ring (associative with 1, but potentially noncommutative). Show that $Z(\mathrm{M}_n(R)) = Z(R)$ and that any two-sided ideal of $\mathrm{M}_n(R)$ is of the form $\mathrm{M}_n(I) \subseteq \mathrm{M}_n(R)$ where I is a two-sided ideal of R.

▹ 6. Let F be a field, let B a simple algebra, and let I be a left B-module with $\dim_F I = \dim_F B$. Show that I is isomorphic to B as a left B-module, i.e., there exists $\alpha \in I$ such $I = B\alpha$.

7. In this exercise, we consider extensions of the Skolem–Noether theorem.

   (a) Let B be a quaternion algebra over F and let $K_1, K_2 \subset B$ be F-subalgebras (not necessarily subfields). Suppose that $\phi\colon K_1 \xrightarrow{\sim} K_2$ is an isomorphism of F-algebras. Show that ϕ lifts to an inner automorphism of B. *[Hint: repeat the proof of Corollary 7.1.5.]*
   (b) Show by example that Corollary 7.7.3 is false if the minimal polynomials are not supposed to be irreducible. In particular, provide an example of isomorphic algebras $K_1, K_2 \subseteq B$ that are not isomorphic by an inner automorphism of B.

8. Let B be a quaternion algebra over F, and let $K \subseteq B$ be a separable, quadratic F-subalgebra. Show that there exists $b \in F^\times$ such that $B \simeq (K \mid b)$. *[Hint: lift the standard involution on K via the Skolem–Noether theorem.]*

▹ 9. Let B be a finite-dimensional F-algebra. Show that if $\alpha \in \operatorname{rad} B$, then $1-\beta\alpha \in B^\times$ for all $\beta \in B$. *[Hint: if $1-\beta\alpha$ is not left invertible then it belongs to a maximal left ideal; left invertible implies invertible.]*

▹ 10. Extend Corollary 7.1.5 to the case where $K = F \times F$ as follows: show that if $K_1, K_2 \subseteq B$ are F-subalgebras with $K_1 \simeq F \times F$, and $\phi\colon K_1 \xrightarrow{\sim} K_2$ is an isomorphism of F-algebras, then ϕ lifts to an inner automorphism of B.

11. Let $n \in \mathbb{Z}_{\geq 2}$ and let F be a field with $\operatorname{char} F \nmid n$. Let $\zeta \in F$ be a primitive nth root of unity. Let $a, b \in F^\times$ and let $A = \left(\dfrac{a, b}{F, \zeta}\right)$ be the algebra over F generated by elements i, j subject to

$$i^n = a, \quad j^n = b, \quad ji = \zeta ij.$$

   (a) Show that $\dim_F A = n^2$.
   (b) Show that A is a central simple algebra over F.
   (c) Let $K = F[i] \simeq F[x]/(x^n - a)$. Show that if $b \in \operatorname{Nm}_{K|F}(K^\times)$ then $A \simeq \mathrm{M}_n(F)$. [Such algebras are called **cyclic algebras** or sometimes **power norm residue algebras**.]

12. Generalize the statement of Proposition 7.5.3(a) as follows. Let A, B be F-algebras, and let $A' \subseteq A$ and $B' \subseteq B$ be F-subalgebras. Prove that

$$C_{A\otimes B}(A' \otimes B') = C_A(A') \otimes C_B(B').$$

13. Let B be a finite-dimensional F-algebra. Show that the following are equivalent:

   (i) B is separable;
   (ii) B is semisimple and the center $K = Z(B)$ is separable;
   (iii) $B \otimes_F B^{\mathrm{op}}$ is semisimple.

14. Let $G \neq \{1\}$ be a finite group. Show that the **augmentation ideal**, the two-sided ideal generated by $g-1$ for $g \in G$, is a nontrivial ideal, and hence $F[G]$ is not simple as an F-algebra.

15. Let G be a finite group of order $n = \#G$. Show that $F[G]$ is a separable F-algebra if and only if $\operatorname{char} F \nmid n$ as follows. [This exercise is known as *Maschke's theorem*.]

    (a) Suppose first that $\operatorname{char} F = 0$ for a special but quick special case. Compute the trace pairing and conclude $F[G]$ is separable.
    (b) If $\operatorname{char} F \mid n$, show that $N = \sum_{g\in G} g$ is a nilpotent element in the center of $F[G]$, so $F[G]$ is not semisimple.
    (c) Suppose that $\operatorname{char} F \nmid n$. Let $B = F[G]$. Define the map of left B-modules by

    $$\phi\colon B \to B \otimes_F B^{\mathrm{op}} =: B^{\mathrm{e}}$$
    $$\phi(1) = \frac{1}{n}\sum_{g\in g} g \otimes g^{-1}$$

    so that $\phi(\alpha) = \alpha\phi(1)$ for all $\alpha \in B$. Give B the structure of a B^{e}-algebra by $(\alpha, \alpha^{\mathrm{o}}) \cdot \beta \mapsto \alpha\beta\alpha^{\mathrm{o}}$. Show that ϕ is a homomorphism of B^{e}-modules, and that the structure map $\psi\colon B^{\mathrm{e}} \to B$ has $\psi \circ \phi = \mathrm{id}_B$. Conclude that B is separable.

16. Let B be an F-algebra, and let F^{al} be an algebraic closure of F. Show that B is simple if and only if $B \otimes_F F^{\mathrm{al}}$ is simple.

17. Let D be a (finite-dimensional) division algebra over F^{al}. Show that $D = F^{\mathrm{al}}$. Conclude that if B is a simple algebra over F^{al}, then $B \simeq \mathrm{M}_n(F^{\mathrm{al}})$ for some $n \geq 1$ and hence is central.

▹ 18. Let B be a (finite-dimensional) F algebra, and let $K \supseteq F$ be a finite separable extension of fields. Show that $\operatorname{rad}(B \otimes_F K) = \operatorname{rad}(B) \otimes_F K$.

▹ 19. Show that if B is a semisimple F-algebra, then so is $\mathrm{M}_n(B)$ for any $n \in \mathbb{Z}_{\geq 1}$.

▹ 20. Let B be a (finite-dimensional) F-algebra with standard involution and suppose $\operatorname{char} F \neq 2$.

    (a) Show that $\operatorname{rad} B = \operatorname{rad}\operatorname{nrd}$. Conclude B is semisimple if and only if $\operatorname{rad}\operatorname{nrd} = \{0\}$.
    (b) Suppose $B \neq F$ and B is central. Conclude that B is a quaternion algebra if and only if $\operatorname{rad}\operatorname{nrd} = \{0\}$.

21. Compute the Jacobson radical $\operatorname{rad} B$ of the F-algebra B with basis $1, i, j, ij$ satisfying

    $$i^2 = a,\ j^2 = 0,\ \text{ and }\ ij = -ji$$

    for $a \in F$, and compute $B/\operatorname{rad} B$. In particular, conclude that such an algebra is not semisimple, so B is not a quaternion algebra. *[Hint: restrict to the case* $\operatorname{char} F \neq 2$ *first.]*

22. Give an example of (finite-dimensional) simple algebras A, B over a field F such that $A \otimes_F B$ is not simple. Then find A, B such that $A \otimes_F B$ is not semisimple.

▹ 23. In Exercise 7.18, we saw that if D is a (finite-dimensional) division algebra over F then $D \otimes_F F^{\mathrm{al}} \simeq \mathrm{M}_n(F^{\mathrm{al}})$ for some $n \geq 1$. In this exercise, we show the same is true if we consider the separable closure. (We proved this already in Exercise 6.3 for D a quaternion algebra.)
Let F be a **separably closed** field, so every nonconstant separable polynomial with coefficients in F has a root in F. Let D be a finite-dimensional central division algebra over F with $\operatorname{char} F = p$. For purposes of contradiction, assume that $D \neq F$.

(a) Prove that $\dim_F D$ is divisible by p.
(b) Show that the minimal polynomial of each nonzero $d \in D$ has the form $x^{p^e} - a$ for some $a \in F$ and $e \geq 0$.
(c) Choose an F^{al}-algebra isomorphism $\phi \colon D \otimes_F F^{\mathrm{al}} \xrightarrow{\sim} \mathrm{M}_n(F^{\mathrm{al}})$. Show that $\operatorname{tr} \phi(x \otimes 1) = 0$ for all $x \in D$.
(d) Prove that D does not exist.

24. Let $K \supseteq F$ be a separable (possibly infinite) extension, and let $f(T) \in K[T]$. Suppose that $f(T)^n \in F[T]$ for some $n \in \mathbb{Z}_{\geq 1}$. Show that $f(T) \in F[T]$. *[Hint: when* $p = \operatorname{char} F \mid n$, *use the fact that* $a^p \in F$ *implies* $a \in F$.*]*

25. Let B be a finite-dimensional F-algebra, let $\alpha \in B$, and let $f_{\mathsf{L}}(\alpha;T)$ and $f_{\mathsf{R}}(\alpha;T)$ be the characteristic polynomial of left and right multiplication of α on B, respectively.

(a) If B is semisimple, show that $f_{\mathsf{L}}(\alpha;T) = f_{\mathsf{R}}(\alpha;T)$.
(b) Give an example where $f_{\mathsf{L}}(\alpha;T) \neq f_{\mathsf{R}}(\alpha;T)$.

▹ 26. Use the Skolem–Noether theorem to give another solution to Exercise 6.2: if $K \subset B$ is a separable quadratic F-algebra then $B \simeq (K, b \mid F)$ for some $b \in F^\times$.

27. Give a direct proof of Corollary 7.7.4. *[Hint: Use the fact that there is a unique simple left B-module.]*

28. Let $B = (K, b \mid F)$ be a quaternion algebra. Show that the subgroup of $\operatorname{Aut}(B)$ that maps $K \subseteq B$ to itself is isomorphic to the group

$$K^\times/F^\times \cup j(K^\times/F^\times).$$

Show that the subgroup of $\operatorname{Aut}(B)$ that restricts to the identity on K (fixing K elementwise) is isomorphic to $K^\times/F^\times$.

▹ 29. Use the Skolem–Noether theorem and the fact that a finite group cannot be written as the union of the conjugates of a proper subgroup to prove Wedderburn's little theorem: a finite division ring is a field.

▹ 30. Let B be a quaternion algebra over F. In this exercise, we show that the commutator subgroup

$$[B^\times, B^\times] = \langle \alpha\beta\alpha^{-1}\beta^{-1} : \alpha, \beta \in B^\times \rangle$$

is precisely

$$[B^\times, B^\times] = B^1 = \{\gamma \in B^\times : \operatorname{nrd}(\gamma) = 1\} = \mathrm{SL}_1(B).$$

(a) Show that $[B^\times, B^\times] \leq B^1$.

(b) Show that $[\mathrm{GL}_2(F), \mathrm{GL}_2(F)] = \mathrm{SL}_2(F)$ if $\#F > 3$. *[Hint: choose $z \in F$ such that $z^2 - 1 \in F^\times$, let $\gamma = \begin{pmatrix} z & 0 \\ 0 & z^{-1} \end{pmatrix}$, and show that for all $x \in F$ we have*

$$\begin{pmatrix} 1 & x \\ 0 & 1 \end{pmatrix} = \left[\gamma, \begin{pmatrix} 1 & x(z^2-1)^{-1} \\ 0 & 1 \end{pmatrix} \right]$$

and analogously for the transpose. See 28.3 for a review of elementary matrices.]

(c) Suppose that B is a division algebra. Let $\gamma \in B^1$. Show that there exists $\alpha \in K = F(\gamma)$ such that $\alpha\overline{\alpha}^{-1} = \gamma$. *[Hint: This is a special case of Hilbert's theorem 90. Let $\alpha = \gamma + 1$ if $\gamma \neq -1$, and $\alpha \in B^0 \setminus \{0\}$ if $\gamma = -1$, with appropriate modifications if* $\operatorname{char} F = 2$.*]* Conclude from the Skolem–Noether theorem that there exists $\beta \in B^\times$ such that $\beta\alpha\beta^{-1} = \overline{\alpha}$, and thus $\gamma \in [B^\times, B^\times]$.

31. Show that every *ring* automorphism of $\mathbb{H}$ is inner. (Compare this with ring automorphisms of $\mathbb{C}$!)

This chapter is licensed under the terms of the Creative Commons Attribution-NonCommercial 4.0 International License (http://creativecommons.org/licenses/by-nc/4.0/), which permits any noncommercial use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license and indicate if changes were made.

The images or other third party material in this chapter are included in the chapter's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the chapter's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.

Chapter 8

Simple algebras and involutions

In this chapter, we examine further connections between quaternion algebras, simple algebras, and involutions.

8.1 ▹ The Brauer group and involutions

An involution on an F-algebra B induces an isomorphism $^{-} : B \xrightarrow{\sim} B^{\mathrm{op}}$, for example such an isomorphism is furnished by the standard involution on a quaternion algebra B. More generally, if B_1, B_2 are quaternion algebras, then the tensor product $B_1 \otimes_F B_2$ has an involution provided by the standard involution on each factor giving an isomorphism to $(B_1 \otimes_F B_2)^{\mathrm{op}} \simeq B_1^{\mathrm{op}} \otimes_F B_2^{\mathrm{op}}$—but this involution is no longer a standard involution (Exercise 8.1). The algebra $B_1 \otimes_F B_2$ is a central simple algebra over F called a **biquaternion algebra**. In some circumstances, we may have

$$B_1 \otimes_F B_2 \simeq \mathrm{M}_2(B_3) \tag{8.1.1}$$

where B_3 is again a quaternion algebra, and in other circumstances, we may not; following Albert, we begin this chapter by studying (8.1.1) and biquaternion algebras in detail.

To this end, we look at the set of isomorphism classes of central simple algebras over F, which is closed under tensor product; if we think that the matrix ring is something that is 'no more complicated than its base ring', it is natural to introduce an equivalence relation on central simple algebras that identifies a division ring with the matrix ring (of any rank) over this division ring. More precisely, if A, A' are central simple algebras over F we say A, A' are **Brauer equivalent** if there exist $n, n' \geq 1$ such that $\mathrm{M}_n(A) \simeq \mathrm{M}_{n'}(A')$. In this way, (8.1.1) reads $B_1 \otimes_F B_2 \sim B_3$. The set of Brauer equivalence classes $[A]$ has the structure of a group under tensor product, known as the **Brauer group** $\mathrm{Br}(F)$ of F, with identity element $[F]$ and inverse $[A]^{-1} = [A^{\mathrm{op}}]$. The class $[B] \in \mathrm{Br}(F)$ of a quaternion algebra B is a 2-torsion element, and therefore so is a biquaternion algebra. In fact, by a striking theorem of Merkurjev, when $\mathrm{char}\, F \neq 2$, all 2-torsion elements in $\mathrm{Br}(F)$ are represented by a tensor product of quaternion algebras (see section 8.3).

© The Author(s) 2021

J. Voight, *Quaternion Algebras*, Graduate Texts in Mathematics 288,
https://doi.org/10.1007/978-3-030-56694-4_8

Finally, our interest in involutions in Chapter 3 began with an observation of Hamilton: the product of a nonzero element with its involute in $\mathbb{H}$ is a positive real number (its norm, or square length). We then proved that the existence of such an involution characterizes quaternion algebras in an essential way. However, one may want to relax this setup and instead consider when the product of a nonzero element with its involute merely has positive *trace*. Such involutions are called **positive** involutions and they arise naturally in algebraic geometry: the Rosati involution is a positive involution on the endomorphism algebra of an abelian variety, and it is a consequence that this algebra (over $\mathbb{Q}$) is semisimple, and unsurprisingly quaternion algebras once again feature prominently (see sections 8.4–8.5).

8.2 Biquaternion algebras

Let F be a field. All tensor products in this section will be taken over F.

8.2.1. Let B_1, B_2 be quaternion algebras over F. The tensor product $B_1 \otimes B_2$ is a central simple algebra over F of dimension $4^2 = 16$ called a **biquaternion algebra**. A biquaternion algebra may be written as a tensor product of two quaternion algebras in different ways, so the pair is not intrinsic to the biquaternion algebra.

By the Wedderburn–Artin theorem (Main Theorem 7.3.10), we have exactly one of the three following possibilities for this algebra:

- $B_1 \otimes B_2$ is a division algebra;
- $B_1 \otimes B_2 \simeq \mathrm{M}_2(B_3)$ where B_3 is a quaternion division algebra over F; or
- $B_1 \otimes B_2 \simeq \mathrm{M}_4(F)$.

We could combine the latter two and just say that $B_1 \otimes B_2 \simeq \mathrm{M}_2(B_3)$ where B_3 is a quaternion algebra over F, since $\mathrm{M}_2(\mathrm{M}_2(F)) \simeq \mathrm{M}_4(F)$ as F-algebras.

Example 8.2.2. By Exercise 8.2, when $\operatorname{char} F \neq 2$ we have

$$\left(\frac{a, b_1}{F}\right) \otimes \left(\frac{a, b_2}{F}\right) \simeq \mathrm{M}_2(B_3)$$

where $B_3 = \left(\frac{a, b_1 b_2}{F}\right)$. In particular, $\left(\frac{a, b}{F}\right) \otimes \left(\frac{a, b}{F}\right) \simeq \mathrm{M}_4(F)$, since $\left(\frac{a, b^2}{F}\right) \simeq \mathrm{M}_2(F)$.

Example 8.2.2 is no accident, as the following proposition indicates.

Proposition 8.2.3. (Albert). *The following are equivalent:*

(i) *There exists a quadratic field extension $K \supset F$ that can be embedded as an F-algebra in both B_1 and B_2;*
(ii) *B_1 and B_2 have a common quadratic splitting field; and*
(iii) *$B_1 \otimes B_2$ is not a division algebra.*

Proof. The equivalence (i) $\Leftrightarrow$ (ii) follows from Lemma 5.4.7.

For the implication (i) $\Rightarrow$ (iii), for $i = 1, 2$ let $\alpha_i \in B_i$ generate K so $\alpha_i^2 = t\alpha_i - n$ with $t, n \in F$. Let

$$\beta := \alpha_1 \otimes 1 - 1 \otimes \alpha_2.$$

Then

$$\begin{aligned} \beta(\alpha_1 \otimes 1 + 1 \otimes \alpha_2 - t) &= \alpha_1^2 \otimes 1 - 1 \otimes \alpha_2^2 - t\beta \\ &= (t\alpha_1 - n) \otimes 1 - 1 \otimes (t\alpha_2 - n) - t\beta = 0. \end{aligned} \tag{8.2.4}$$

Therefore β is a zerodivisor and $B_1 \otimes B_2$ is not a division algebra.
To finish, we prove (iii) $\Rightarrow$ (i). We have an embedding

$$\begin{aligned} B_1 &\hookrightarrow B_1 \otimes B_2 \\ \alpha &\mapsto \alpha \otimes 1 \end{aligned}$$

and similarly B_2; the images of B_1 and B_2 in $B_1 \otimes B_2$ commute. Write $B_2 = (K, b_2 \mid F)$. Consider $(B_1)_K = B_1 \otimes K \subset B_1 \otimes B_2$; then $(B_1)_K$ is a quaternion algebra over K (with $\dim_F (B_1)_K = 8$). If $(B_1)_K$ is not a division algebra, then K splits B_1 and $K \hookrightarrow B_1$ and we are done. So suppose that $(B_1)_K$ is a division algebra. Then

$$B_1 \otimes B_2 = (B_1)_K + (B_1)_K j$$

is free of rank 2 as a left $(B_1)_K$-module.

Since $B_1 \otimes B_2 \simeq \mathrm{M}_2(B_3)$ is not a division algebra, there exists $\epsilon \in B_1 \otimes B_2$ nonzero such that $\epsilon^2 = 0$. Without loss of generality, we can write $\epsilon = \alpha_1 \otimes z + j$ where $\alpha_1 \in B_1$ and $z \in K$. Then

$$0 = \epsilon^2 = \alpha_1^2 \otimes z^2 + (\alpha_1 \otimes z)j + (\alpha_1 \otimes \overline{z})j + b_2. \tag{8.2.5}$$

From the basis $1, j$ over $(B_1)_K$, if $\overline{z} = t - z$ with $t \in F$, we conclude that

$$\alpha_1 \otimes z + \alpha_1 \otimes (t - z) = \alpha_1 \otimes t = 0.$$

Therefore $t = 0$, and $z^2 = c$ for some $c \in F^\times$. Then from (8.2.5) $c\alpha_1^2 + b_2 = 0$ so $\alpha_1^2 = -b_2/c$ and B_1 contains the quadratic field $F(\sqrt{-b_2 c})$. But so does B_2, as $(zj)^2 = -b_2 c$ as well.

(For an alternate proof, see Jacobson [Jacn2009, Theorem 2.10.3].) $\square$

Remark 8.2.6. In view of Proposition 8.2.3, we say that two quaternion algebras B_1, B_2 over F are **linked** if they contain a common quadratic field extension $K \supseteq F$. For further discussion of biquaternion algebras and linkage in characteristic 2 (where one must treat separable and inseparable extensions differently), see Knus [Knu93], Lam [Lam2002], or Sah [Sah72]. Garibaldi–Saltman [GS2010] study the subfields of quaternion algebra over fields with char $F \neq 2$.

From now on, we suppose that char $F \neq 2$. (For the case char $F = 2$, see Chapman–Dolphin–Laghbribi [CDL2015, §6].)

8.2.7. Motivated by Proposition 8.2.3, we consider the quadratic extensions represented by B_1 and B_2 encoded in the language of quadratic forms (recalling Lemma 5.5.4). Let

$$V = \{\alpha_1 \otimes 1 - 1 \otimes \alpha_2 \in B_1 \otimes B_2 : \operatorname{trd}(\alpha_1) = \operatorname{trd}(\alpha_2)\}.$$

Then $\dim_F V = 6$, and we may identify $V = B_1^0 \otimes 1 - 1 \otimes B_2^0$. The reduced norm on each factor separately defines a quadratic form on V by taking the difference: explicitly, if $B_1 = (a_1, b_1 \mid F)$ and $B_2 = (a_2, b_2 \mid F)$, then taking the standard bases for B_1, B_2

$$\begin{aligned} Q(B_1, B_2) &\simeq \langle -a_1, -b_1, a_1 b_1\rangle \boxplus -\langle -a_2, -b_2, a_2 b_2\rangle \\ &\simeq \langle -a_1, -b_1, a_1 b_1, a_2, b_2, -a_2 b_2\rangle. \end{aligned}$$

The quadratic form $Q(B_1, B_2) : V \to F$ is called the **Albert form** of the biquaternion algebra $B_1 \otimes B_2$.

We then add onto Proposition 8.2.3 as follows.

Proposition 8.2.8. (Albert). *Let $B_1 \otimes B_2$ be a biquaternion algebra over F (with* $\operatorname{char} F \neq 2$*) with Albert form $Q(B_1, B_2)$. Then the following are equivalent:*

(i) *B_1, B_2 have a common quadratic splitting field;*
(iv) *$Q(B_1, B_2)$ is isotropic.*

Proof. The implication (ii) $\Rightarrow$ (iv) follows by construction 8.2.7. To prove (iv) $\Rightarrow$ (ii), without loss of generality, we may suppose B_1, B_2 are division algebras; then an isotropic vector of Q corresponds to elements $\alpha_1 \in B_1$ and $\alpha_2 \in B_2$ such that $\alpha_1^2 = \alpha_2^2 = c \in F^\times$. Therefore $K = F(\sqrt{c})$ is a common quadratic splitting field. □

Remark 8.2.9. Albert's book [Alb39] on algebras still reads well today. The proof of Proposition 8.2.3 is due to him [Alb72]. ("I discovered this theorem some time ago. There appears to be some continuing interest in it, and I am therefore publishing it now.") Albert used Proposition 8.2.8 to show that

$$B_1 = \left(\frac{-1,-1}{F}\right) \quad \text{and} \quad B_2 = \left(\frac{x,y}{F}\right)$$

over $F = \mathbb{R}(x, y)$ have tensor product $B_1 \otimes_F B_2$ a division algebra by verifying that the Albert form $Q(B_1, B_2)$ is anisotropic over F. See Lam [Lam2005, Example VI.1] for more details. For the fields of interest in this book (local fields and global fields), a biquaternion algebra will never be a division algebra—the proof of this fact rests on classification results for quaternion algebras over these fields, which we will take up in earnest in Part II.

8.3 Brauer group

Motivated to study the situation where $B_1 \otimes B_2 \simeq M_2(B_3)$ among quaternion algebras B_1, B_2, B_3 more generally, we now turn to the Brauer group.

Let $\operatorname{CSA}(F)$ be the set of isomorphism classes of central simple F-algebras. The operation of tensor product on $\operatorname{CSA}(F)$ defines a commutative binary operation with

identity F, but inverses are lacking (for dimension reasons). So we define an equivalence relation $\sim$ on $\mathrm{CSA}(F)$ by

$$A \sim A' \text{ if } \mathrm{M}_{n'}(A) \simeq \mathrm{M}_n(A') \text{ for some } n, n' \geq 1 \tag{8.3.1}$$

and we say then that A, A' are **Brauer equivalent**. In particular, $A \sim \mathrm{M}_n(A)$ for all $A \in \mathrm{CSA}(F)$ as needed above.

Lemma 8.3.2. *The set of equivalence classes of central simple F-algebras under the equivalence relation $\sim$ has the structure of an abelian group under tensor product, with identity $[F]$ and inverse $[A]^{-1} = [A^{\mathrm{op}}]$.*

Proof. By Exercise 8.5, the operation is well-defined: if $A, A' \in \mathrm{CSA}(F)$ and $A' \sim A'' \in \mathrm{CSA}(F)$ then $A \otimes A' \sim A \otimes A''$. To conclude, we need to show that inverses exist. This is furnished by Lemma 7.5.4: if $\dim_F A = n$ and A^{op} is the opposite algebra of A (3.2.2) then the map

$$\begin{aligned} A \otimes_F A^{\mathrm{op}} &\to \mathrm{End}_F(A) \simeq \mathrm{M}_n(F) \\ \alpha \otimes \beta &\mapsto (\mu \mapsto \alpha\mu\beta) \end{aligned}$$

is an isomorphism of F-algebras, so $[A]^{-1} = [A^{\mathrm{op}}]$ provides an inverse to $[A]$. □

So we make the following definition.

Definition 8.3.3. The **Brauer group** of F is the set $\mathrm{Br}(F)$ of Brauer equivalence classes of central simple F-algebras (8.3.1) under the group operation of tensor product.

8.3.4. Let B be a quaternion algebra over F. We have $B \simeq \mathrm{M}_2(F)$ if and only if $[B] = [F]$ is the identity. Otherwise, B is a division algebra. Then the standard involution gives an F-algebra isomorphism $B \xrightarrow{\sim} B^{\mathrm{op}}$, and hence in $\mathrm{Br}(F)$ we have $[B]^{-1} = [B]$ and so $[B]$ is an element of order 2. Since $\mathrm{Br}(F)$ is abelian, it follows that biquaternion algebras, or more generally tensor products $B_1 \otimes \cdots \otimes B_t$ of quaternion algebras B_i, are also elements of order at most 2 in $\mathrm{Br}(F)$.

Theorem 8.3.5. (Merkurjev). *Let* $\operatorname{char} F \neq 2$. *Then* $\mathrm{Br}(F)[2]$ *is generated by quaternion algebras over F, i.e., every (finite-dimensional) central division F-algebra with involution is Brauer equivalent to a tensor product of quaternion algebras.*

Remark 8.3.6. More generally, Merkurjev [Mer82] proved in 1981 that a division algebra with an involution is Brauer equivalent to a tensor product of quaternion algebras; more precisely, if D is a division F-algebra with (not necessarily standard) involution, then there exists $n \in \mathbb{Z}_{\geq 1}$ such that $\mathrm{M}_n(D)$ is isomorphic to a tensor product of quaternion algebras. His theorem, more properly, says that the natural map $K_2(F) \to \mathrm{Br}(F)[2]$ is an isomorphism. (Some care is required in this area: for example, Amitsur–Rowen–Tignol [ART79] exhibit a division algebra D of degree 8 with involution that is not a tensor product of quaternion algebras, but $\mathrm{M}_2(D)$ is a tensor product of quaternion algebras.) For an elementary proof of Merkurjev's theorem, see Wadsworth [Wad86].

Remark 8.3.7. Just as quaternion algebras are in correspondence with conics (Corollary 5.5.2), with a quaternion algebra split if and only if the corresponding conic has a rational point (Theorem 5.5.3), similarly the Brauer group of a field has a geometric interpretation (see e.g. Serre [Ser79, §X.6]): central simple algebras correspond to *Brauer–Severi varieties*—for each degree $n \geq 1$, both are parametrized by the Galois cohomology set $H^1(\mathrm{Gal}(F^{\mathrm{sep}}\,|\,F), \mathrm{PGL}_n)$.

8.4 Positive involutions

We now turn to study algebras with involution more general than a standard involution. Throughout this section, let $F \subseteq \mathbb{R}$ be a subfield of $\mathbb{R}$ and B a finite-dimensional F-algebra. We define the trace map $\mathrm{Tr} : B \to \mathbb{R}$ by the trace of left multiplication.

Definition 8.4.1. An involution $^* : B \to B$ is **positive** if $\mathrm{Tr}(\alpha^*\alpha) > 0$ for all $\alpha \in B \smallsetminus \{0\}$.

Since the map $(\alpha, \beta) \mapsto \mathrm{Tr}(\alpha^*\beta)$ is bilinear, an involution * on B is positive if and only if $\mathrm{Tr}(\alpha^*\alpha) > 0$ for α in a basis for B and so is positive if and only if its extension to $B \otimes_F \mathbb{R}$ is positive.

Example 8.4.2. The standard involutions on $\mathbb{R}$, $\mathbb{C}$, and $\mathbb{H}$, defined by $\alpha \mapsto \mathrm{trd}(\alpha) - \alpha$, are positive involutions. The standard involution on $\mathbb{R} \times \mathbb{R}$ is not positive since for $\alpha = (x_1, x_2) \in \mathbb{R} \times \mathbb{R}$ we have $\mathrm{Tr}(\alpha\overline{\alpha}) = 2x_1x_2$. The standard involution on $\mathrm{M}_2(\mathbb{R})$ is also not positive, since for $\alpha \in \mathrm{M}_2(\mathbb{R})$ we have $\mathrm{Tr}(\alpha\overline{\alpha}) = 4\det(\alpha)$.

8.4.3. Let D be one of $\mathbb{R}$, $\mathbb{C}$, or $\mathbb{H}$. Let $B = \mathrm{M}_n(D)$. The standard involution $^-$ on D extends to an involution on B, acting on coordinates. The **conjugate transpose** (or, perhaps better the **standard involution transpose**) map

$$\begin{aligned} ^* : B &\to B \\ \alpha &\mapsto \alpha^* = \overline{\alpha}^{\mathsf{t}} \end{aligned}$$

also defines an involution on B, where $^{\mathsf{t}}$ is the transpose map. If $\alpha = (a_{ij})_{i,j=1,\dots,n}$ then

$$\mathrm{Tr}(\alpha^*\alpha) = n(\dim_{\mathbb{R}} D) \sum_{i,j=1}^{n} \overline{a_{ij}}a_{ij} > 0; \tag{8.4.4}$$

thus * is positive, and the norm $\alpha \mapsto \mathrm{Tr}(\alpha^*\alpha)$ is (an integer multiple of) the **Frobenius norm** on B.

We will soon see that every positive involution can be derived from the conjugate transpose as in 8.4.3. First, we reduce to the case where B is a semisimple algebra.

Lemma 8.4.5. *Suppose that B admits a positive involution *. Then B is semisimple.*

Proof. We give two proofs. First, we appeal to Theorem 7.9.4: since the trace pairing is positive definite, it is nondegenerate and immediately B is semisimple.

For a second (more general) proof, let $J = \operatorname{rad} B$ be the Jacobson radical of B. By Lemma 7.4.2, B is semisimple if and only if $\operatorname{rad} B = \{0\}$, and by Lemma 7.4.8, $J = \operatorname{rad} B$ is nilpotent. Suppose for purposes of contradiction that $J \neq \{0\}$. Then there exists $n > 0$ such that $J^n \neq \{0\}$ but $J^{n+1} = \{0\}$. Let $\epsilon \in J$ be such that $\epsilon^n \neq 0$ but $\epsilon^{n+1} = 0$. The involution gives an isomorphism $B \to B^{\mathrm{op}}$ taking maximal left ideals to maximal right ideals and therefore by Corollary 7.4.6 we conclude $J^* = J$. Thus $\epsilon^n\epsilon^* = 0$ so $\operatorname{Tr}(\epsilon^n(\epsilon^*)^n) = \operatorname{Tr}(\epsilon^n(\epsilon^n)^*) = 0$, contradicting that * is positive. □

8.4.6. Suppose B is semisimple with a positive involution *, and let B_i be a simple factor of B. Then * preserves B_i: for if $B_i^* = B_j \neq B_i$, then B_j is a simple factor and $B_iB_j = 0$ so $\operatorname{Tr}(B_iB_i^*) = \operatorname{Tr}(B_iB_j) = \{0\}$, a contradiction.

Putting Lemma 8.4.5 with 8.4.6, we see it is enough to classify positive involutions on simple $\mathbb{R}$-algebras. By the theorem of Frobenius (Corollary 3.5.8), a simple algebra over $\mathbb{R}$ is isomorphic to $\mathrm{M}_n(D)$ with $D = \mathbb{R}, \mathbb{C}, \mathbb{H}$, so 8.4.3 applies.

Proposition 8.4.7. *Let $B \simeq \mathrm{M}_n(D)$ be a simple $\mathbb{R}$-algebra and let * be the conjugate transpose involution on B. Let $^\dagger \colon B \to B$ be another positive involution on B. Then there exists an element $\mu \in B^\times$ with $\mu^* = \mu$ such that*

$$\alpha^\dagger = \mu^{-1}\alpha^*\mu$$

for all $\alpha \in B$.

Proof. First suppose B is central over $\mathbb{R}$. Then the involutions † and * give two $\mathbb{R}$-algebra maps $B \to B^{\mathrm{op}}$. By the Skolem–Noether theorem (Main Theorem 7.7.1), there exists $\mu \in B^\times$ such that $\alpha^\dagger = \mu^{-1}\alpha^*\mu$. Since

$$\alpha = (\alpha^\dagger)^\dagger = (\mu^{-1}\alpha^*\mu)^\dagger = \mu^{-1}(\mu^{-1}\alpha^*\mu)^*\mu = (\mu^{-1}\mu^*)\alpha(\mu^{-1}\mu^*)^{-1} \tag{8.4.8}$$

for all $\alpha \in B$, we have $\mu^{-1}\mu^* \in Z(B) = \mathbb{R}$, so $\mu^* = c\mu$ for some $c \in \mathbb{R}$. But $(\mu^*)^* = \mu = (c\mu^*)^* = c^2\mu$, thus $c = \pm 1$. But if $c = -1$, then μ is skew-symmetric so its top-left entry is $\mu_{11} = 0$; but then for the matrix unit e_{11} we have

$$\operatorname{Tr}(e_{11}e_{11}^\dagger) = \operatorname{Tr}(e_{11}\mu^{-1}e_{11}^*\mu) = \operatorname{Tr}(\mu^{-1}e_{11}\mu e_{11}) = \operatorname{Tr}(\mu^{-1}\mu_{11}) = 0, \tag{8.4.9}$$

a contradiction.

A similar argument holds if B has center $Z(B) = \mathbb{C}$. The restriction of an involution to $Z(B)$ is either the identity or complex conjugation; the latter holds for the conjugate transpose involution, as well as for †: if $z \in Z(B)$ then $\operatorname{Tr}(zz^\dagger) = n^2(zz^\dagger) > 0$, and we must have $z^\dagger = \overline{z}$. So the map $\alpha \mapsto (\alpha^*)^\dagger$ is a $\mathbb{C}$-linear automorphism, and again there exists $\mu \in B^\times$ such that $\alpha^\dagger = \mu^{-1}\alpha^*\mu$. By the same argument, we have $\mu^* = z\mu$ with $z \in \mathbb{C}$, but now $\mu = (\mu^*)^* = \overline{z}z\mu$ so $|z| = 1$. Let $w^2 = w/\overline{w} = z$; then $(w\mu)^* = \overline{w}\mu^* = \overline{w}z\mu = w\mu$. Replacing μ by $w\mu$, we may take $z = 1$. □

Corollary 8.4.10. *The only positive involution on a real division algebra is the standard involution.*

Proof. Apply Proposition 8.4.7 with $n = 1$, noting that $\mu^* = \overline{\mu} = \mu$ implies $\mu \in \mathbb{R}$. □

8.4.11. Let $\mu \in B^\times$ with $\mu^* = \mu$. Then μ is self-adjoint with respect to the pairing $(\alpha, \beta) \mapsto \operatorname{Tr}(\alpha^*\beta)$:

$$(\mu\alpha, \beta) = \operatorname{Tr}((\mu\alpha)^*\beta) = \operatorname{Tr}(\alpha^*\mu^*\beta) = \operatorname{Tr}(\alpha^*\mu\beta) = (\alpha, \mu\beta).$$

It follows from the spectral theorem that the $\mathbb{R}$-linear endomorphism of B given by left-multiplication by μ on B as an $\mathbb{R}$-algebra is diagonalizable (with real eigenvalues) via a symmetric matrix. We say μ is **positive definite** (for *) if all eigenvalues of μ are positive. The map $\alpha \mapsto \operatorname{Tr}(\alpha^*\mu\alpha)$ defines a quadratic form on B, and μ is positive definite if and only this quadratic form is positive definite.

Lemma 8.4.12. *Let $\mu^* = \mu$. Then the involution $\alpha^\dagger = \mu^{-1}\alpha^*\mu$ is positive if and only if either μ or $-\mu$ is positive definite.*

Proof. Diagonalize the quadratic form $\alpha \mapsto \operatorname{Tr}(\alpha^*\mu\alpha)$ to get $\langle a_1, \dots, a_m\rangle$ in a normalized basis $e_1, \dots, e_m$, and suppose without loss of generality that $a_i = \pm 1$. If all $a_i = -1$, then we can replace μ with $-\mu$ without changing the involution to suppose they are all $+1$.

Suppose μ is not positive, and without loss of generality $a_1 < 0$ and $a_2 > 0$, then $\operatorname{Tr}((e_1 + e_2)^*\mu(e_1 + e_2)) = -1 + 1 = 0$, a contradiction. Conversely, if μ is positive definite, then all eigenvalues are $+1$. Let $\nu = \sqrt{\mu}$ be such that $\nu^* = \nu$, and then

$$\begin{aligned} \operatorname{Tr}(\alpha^*\mu^{-1}\alpha\mu) &= \operatorname{Tr}(\alpha^*\nu^{-2}\alpha\nu^2) = \operatorname{Tr}((\nu\alpha^*\nu^{-1})(\nu^{-1}\alpha\nu)) \\ &= \operatorname{Tr}((\nu^{-1}\alpha\nu)^*(\nu^{-1}\alpha\nu)) > 0 \end{aligned} \tag{8.4.13}$$

for all $\alpha \in B$, so † is positive. □

Example 8.4.14. If $n = 1$, and $B = D$, then the condition $\mu^* = \mu$ implies $\mu \in \mathbb{R}$, and the condition μ positive implies $\mu > 0$; rescaling does not affect the involution, so we can take $\mu = 1$ and there is a unique positive involution on D given by *.

Example 8.4.15. Let $B = \mathrm{M}_2(\mathbb{R})$. Then $\mu = \begin{pmatrix} 2a & b \\ b & 2c \end{pmatrix}$ is positive definite if and only if $a > 0$ and $b^2 - 4ac < 0$. Combining Proposition 8.4.7 with Lemma 8.4.12, we see that all positive involutions † on B are given by $\alpha^\dagger = \mu^{-1}\alpha^*\mu$ where μ is positive definite.

We can instead relate positive involutions to the standard involution $\overline{\alpha}$ instead of the transpose; to this end, it is enough to find $j \in B^\times = \mathrm{GL}_2(\mathbb{R})$ such that $\overline{\alpha} = j^{-1}\alpha^* j$, and the element $j = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}$ does the trick, because

$$\begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix} \begin{pmatrix} a & c \\ b & d \end{pmatrix} \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix} = \begin{pmatrix} d & -b \\ -c & a \end{pmatrix}. \tag{8.4.16}$$

From the product $\begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix} \begin{pmatrix} 2a & b \\ b & 2c \end{pmatrix} = \begin{pmatrix} b & 2c \\ -2a & -b \end{pmatrix}$, we conclude that all positive involutions are given by $\alpha^\dagger = \mu^{-1}\overline{\alpha}\mu$ where $\mu^2 \in \mathbb{R}_{<0}$.

Remark 8.4.17. Beyond the application to endomorphism algebras, Weil [Weil60] has given a more general point of view on positive involutions, connecting them to the classical groups. For more on involutions on finite-dimensional algebras over real closed fields, see work of Munn [Mun2004].

8.5 * Endomorphism algebras of abelian varieties

We conclude this chapter with an advanced (optional) application: we characterize endomorphism algebras of (simple) abelian varieties in terms of algebras with involutions. We borrow from the future the notions from section 43.4. Briefly, a *complex torus* of dimension g is a complex manifold of the form $A = V/\Lambda$ for $g \geq 0$, where $\Lambda \subset V \simeq \mathbb{C}^g$ is a lattice (discrete subgroup) and $\Lambda \simeq \mathbb{Z}^{2g}$. A *complex abelian variety* is a certain kind of complex torus. A complex abelian variety A is **simple** if A has no abelian subvariety other than $\{0\}$ and A.

An **endomorphism** of A is a $\mathbb{C}$-linear map $\alpha\colon V \to V$ such that $\alpha(\Lambda) \subseteq \Lambda$. Let $\operatorname{End}(A)$ be the ring ($\mathbb{Z}$-algebra) of endomorphisms of A.

Proposition 8.5.1. $B = \operatorname{End}(A) \otimes \mathbb{Q}$ *is a finite-dimensional algebra over* $\mathbb{Q}$ *that admits a positive involution* ${}^\dagger\colon B \to B$.

Proof. The algebra B acts faithfully on $\Lambda \otimes \mathbb{Q} \simeq \mathbb{Q}^{2g}$, so is isomorphic to a subalgebra of $\mathrm{M}_{2g}(\mathbb{Q})$ hence is finite-dimensional over $\mathbb{Q}$. For positivity, see Proposition 43.4.24 (for the case when A is principally polarized). □

Remark 8.5.2. The involution ${}^\dagger\colon B \to B$ is called the **Rosati involution** (and depends on a choice of **polarization** $\lambda\colon A \to A^\vee$, where $A^\vee$ is the dual abelian variety).

Now Lemma 8.4.5 and Proposition 8.5.1 imply that B is semisimple as a $\mathbb{Q}$-algebra, with

$$B \simeq \prod_{i=1}^{r} \mathrm{M}_{n_i}(D_i)$$

where each $D_i \subseteq B$ is a division algebra. It follows that A is isogenous to a product

$$A_1^{n_1} \times \cdots \times A_r^{n_r}$$

where $n_1, \dots, n_r > 0$ and $A_1, \dots, A_r$ are simple pairwise nonisogenous abelian subvarieties of A such that $D_i = \operatorname{End}(A_i) \otimes_{\mathbb{Z}} \mathbb{Q}$.

We therefore reduce to the case where A is simple, and $D := \operatorname{End}(A) \otimes \mathbb{Q}$ is a division algebra. Let $K := Z(D)$ be the center of D and let

$$K_0 := K^{\langle \dagger \rangle} = \{a \in K : a^\dagger = a\}$$

be the subfield of K where † acts by the identity.

Lemma 8.5.3. K_0 *is a totally real number field, i.e., every embedding* $K_0 \hookrightarrow \mathbb{C}$ *factors through* $\mathbb{R}$*, and if* $\dagger$ *acts nontrivially on* K*, then* K *is a CM field, i.e.,* K *is a totally imaginary extension of* K_0.

Proof. The positive involution † restricts to complex conjugation on $Z(D)$ by Proposition 8.4.7, so for all embeddings $K_0 \hookrightarrow \mathbb{C}$, the image lies in $\mathbb{R}$. For the same reason, we cannot have † acting nontrivially on K and have an embedding $K \hookrightarrow \mathbb{R}$. □

The following theorem of Albert classifies the possibilities for D.

Theorem 8.5.4. (Albert). *Let D be a (finite-dimensional) division algebra over $\mathbb{Q}$ with positive involution † and center $K = Z(D)$, let $K_0 := K^{\langle\dagger\rangle}$ and $n := [K_0 : \mathbb{Q}]$. Then K_0 is a totally real number field, and one of the four following possibilities holds:*

(I) *$D = K = K_0$ and † is the identity;*

(II) *$K = K_0$ and D is a quaternion algebra over K_0 such that*

$$D \otimes_{\mathbb{Q}} \mathbb{R} \simeq \mathrm{M}_2(\mathbb{R})^n,$$

and there exists $\mu \in D^\times$ such that $\mu^2 = d \in K_0^\times$ is totally negative and $\alpha^\dagger = \mu^{-1}\overline{\alpha}\mu$ for all $\alpha \in D$;

(III) *$K = K_0$ and D is a quaternion algebra over K_0 such that*

$$D \otimes_{\mathbb{Q}} \mathbb{R} \simeq \mathbb{H}^n,$$

and † is the standard involution; or

(IV) *$K \supsetneq K_0$ and*

$$D \otimes_{\mathbb{Q}} \mathbb{R} \simeq \mathrm{M}_d(\mathbb{C})^n$$

for some $d \geq 1$, and † extends to the conjugate transpose * on each factor $\mathrm{M}_d(\mathbb{C})$.

Proof. We have assembled many of the tools needed to prove this theorem, and hopefully motivated its statement sufficiently well—but unfortunately, a proof remains just out of reach: we require some results about quaternion algebras over number fields not yet in our grasp. For a proof, see Mumford [Mum70, Application I, §21] or Birkenhake–Lange [BL2004, §§5.3–5.5].

To connect a few dots as well as we can right now, we give a sketch in the case where $K = K_0$ for the reader who is willing to flip ahead to Chapter 14. In this case, D is a central division algebra over $K = K_0$ and has a K_0-linear involution giving an isomorphism $D \xrightarrow{\sim} D^{\mathrm{op}}$ of K_0-algebras. Looking in the Brauer group $\mathrm{Br}(K_0)$, we conclude that $[D] = [D^{\mathrm{op}}] = [D]^{-1}$, so $[D] \in \mathrm{Br}(K_0)$ has order at most 2. By class field theory (see Remark 14.6.10), we conclude that either $D = K_0$ or D is a (division) quaternion algebra over K_0. If $D = K_0$, we are in case (I), so suppose D is a quaternion algebra over K_0. We have $D \otimes_{\mathbb{Q}} \mathbb{R} \simeq \prod_{v|\infty} D_v$ a direct product of n quaternion algebras D_v over $\mathbb{R}$ indexed by the real places v of K_0. We have $D_v \simeq \mathrm{M}_2(\mathbb{R})$ or $D_v \simeq \mathbb{H}$, and our positive involution induces a corresponding positive involution on each D_v. If there exists v such that $D_v \simeq \mathbb{H}$, then by Corollary 8.4.10, the positive involution on D_v is the standard involution, so it is so on D, and then all components must have $D_v \simeq \mathbb{H}$ as the standard involution is not positive on $\mathrm{M}_2(\mathbb{R})$—and we are in case (II). Otherwise, we are in case (III), with Proposition 8.4.7 and Example 8.4.15 characterizing the positive involution. □

Exercises

Let F be a field.

1. Let B_1, B_2 be quaternion algebras over F, with standard involution written $\overline{}$ in both cases. Let $A := B_1 \otimes B_2$.

   (a) Show that the map $\sigma\colon A \to A$ defined by $\alpha_1 \otimes \alpha_2 \mapsto \overline{\alpha_1} \otimes \overline{\alpha_2}$ for $\alpha_1 \in B_1$ and $\alpha_2 \in B_2$ extends to an involution on A, but it is not a standard involution. *[Hint: consider sums.]*

   (b) Suppose that $\operatorname{char} F \neq 2$. Diagonalize $A = A^+ \oplus A^-$ into $+1$ and -1 eigenspaces for σ. Show that
   $$A^+ = F \oplus (B_1^- \otimes B_2^-) \quad \text{and} \quad A^- = (B_1^- \otimes F) \oplus (F \otimes B_2^-).$$

▷ 2. Suppose $\operatorname{char} F \neq 2$ and let $B_1 := \left(\dfrac{a, b_1}{F}\right)$ and $B_2 := \left(\dfrac{a, b_2}{F}\right)$ be quaternion algebras over F.

   (a) Let B_3 be the F-span of $1, i_3 := i_1 \otimes 1, j_3 := j_1 \otimes j_2$, and $k_3 := i_3 j_3 = i_1 j_1 \otimes j_2$ inside $B_1 \otimes B_2$. Show that $B_3 \simeq \left(\dfrac{a, b_1 b_2}{F}\right)$ as F-algebras.

   (b) Similarly, let B_4 be the F-span of 1, $i_4 := 1 \otimes j_2$, $j_4 := (i_1 \otimes k_2)/a$, and $k_4 := i_4 j_4$. Show that $B_4 \simeq \left(\dfrac{b_2, -b_2}{F}\right) \simeq \mathrm{M}_2(F)$.

   (c) Show that
   $$B_1 \otimes B_2 \simeq B_3 \otimes B_4 \simeq \mathrm{M}_2(B_3).$$
   [Hint: Show that B_3 and B_4 *are commuting subalgebras, or consider the map* $B_3 \otimes B_4 \to B_1 \otimes B_2$ given by multiplication.]

   (d) Restore symmetry and repeat (a)–(c) to find algebras $B_3' \simeq B_3$ and $B_4' \simeq \left(\dfrac{b_1, -b_1}{F}\right)$ with $B_1 \otimes B_2 \simeq B_3' \otimes B_4' \simeq \mathrm{M}_2(B_3')$.

3. Suppose $\operatorname{char} F \neq 2$. Show that $B_1 \otimes B_2 \simeq \mathrm{M}_4(F)$ if and only if the Albert form $Q(B_1, B_2)$ is totally hyperbolic.

4. Let G be a finite group. Show that the map induced by $g \mapsto g^{-1}$ for $g \in G$ defines an positive involution on $\mathbb{R}[G]$. Then show that this map composed with coordinatewise complex conjugation defines a positive involution on $\mathbb{C}[G]$ (as an $\mathbb{R}$-algebra).

▷ 5. Show that if $\sim$ is the equivalence relation (8.3.1) on $\mathrm{CSA}(F)$, then $\sim$ is compatible with tensor product, i.e., if $A, A' \in \mathrm{CSA}(F)$ and $A' \sim A'' \in \mathrm{CSA}(F)$ then $A \otimes A' \sim A \otimes A''$.

6. Show that every class in the Brauer group $\mathrm{Br}(F)$ contains a unique division F-algebra, up to isomorphism.

7. Show that $\mathrm{Br}\, F = \{1\}$ if F is separably closed, and that $\mathrm{Br}(\mathbb{R}) \simeq \mathbb{Z}/2\mathbb{Z}$ and $\mathrm{Br}(\mathbb{F}_q) = \{1\}$.

8. Let $B \in \mathrm{CSA}(F)$ and suppose that B has an involution (not necessarily standard). Show that $[B]$ has order at most 2 in $\mathrm{Br}\, F$.
9. Let $K \supseteq F$ be a field extension. Show that the map $A \mapsto A \otimes_F K$ induces a group homomorphism $\mathrm{Br}\, F \to \mathrm{Br}\, K$. Conclude that the set of isomorphism classes of central division F-algebras D such that $D \otimes_F K \simeq \mathrm{M}_n(K)$ for some $n \geq 1$ forms a subgroup of $\mathrm{Br}\, F$, called the **relative Brauer group** $\mathrm{Br}(K \mid F)$.
10. In this exercise, we give an example of a central simple algebra of infinite dimension, called the *Weyl algebra*.

    Suppose $\mathrm{char}\, F = 0$, let $F[x]$ be the polynomial ring over F in the variable x. Inside the enormous algebra $\mathrm{End}_F\, F[x]$ is the operator $f(x) \mapsto xf(x)$, denoted also x, and the differentiation operator $\delta : F[x] \to F[x]$. These two operators are related by the product rule:

    $$\delta(xf(x)) - x\delta(f(x)) = f(x).$$

    Accordingly, the subalgebra of $\mathrm{End}_F\, F[x]$ generated by δ, x is isomorphic to an algebra given in terms of generators and relations:

    $$W := F\langle \delta, x\rangle / \langle \delta x - x\delta - 1\rangle,$$

    the quotient of the "noncommutative polynomial ring" in two variables $F\langle \delta, x\rangle$ by the two-sided ideal generated by $\delta x - x\delta - 1$.

    (a) Show that every element of W can be written in the form $\sum_{i=0}^{n} f_i(x)\delta^i$ where $f_i(x) \in F[x]$ for all i, i.e., W has F-basis elements $x^i\delta^j$ for $i, j \geq 0$.
    (b) Show that $Z(W) = F$.
    (c) Let I be a two-sided of W. Show that if there exists nonzero $f(x) \in F[x] \cap I$, then $I = W$. Similarly, show that if $\delta^n \in I$ for some $n \geq 0$, then $I = W$.
    (d) Show that W is simple. *[Hint: argue by induction.]*

11. Let B be a finite-dimensional $\mathbb{R}$-algebra with positive involution ${}^*\colon B \to B$. Let

    $$P(B, {}^*) := \{\mu \in B : \mu^* = \mu \text{ and } \mu \text{ is positive definite for } {}^*\}.$$

    (a) Show that $B^\times$ acts on $P(B, {}^*)$ by $\beta \cdot \mu := \beta^* \mu \beta$.
    (b) Show that $P(B, {}^*)$ is a convex open subset of $\{\alpha \in B : \alpha^* = \alpha\}$, an $\mathbb{R}$-vector subspace of B.
    (c) Let $\psi\colon B \to B$ be an $\mathbb{R}$-algebra automorphism or anti-automorphism. Show that $\alpha^\dagger := \psi^{-1}(\psi(\alpha)^*)$ defines a positive involution for $\alpha \in B$, and that ψ maps $P(B, {}^\dagger)$ bijectively to $P(B, {}^\psi)$.

This chapter is licensed under the terms of the Creative Commons Attribution-NonCommercial 4.0 International License (http://creativecommons.org/licenses/by-nc/4.0/), which permits any noncommercial use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license and indicate if changes were made.

The images or other third party material in this chapter are included in the chapter's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the chapter's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.

Part II
Arithmetic

Chapter 9

Lattices and integral quadratic forms

In many ways, quaternion algebras are like "noncommutative quadratic field extensions": this is apparent from their very definition, but also from their description as wannabe 2×2-matrices. Just as the quadratic fields $\mathbb{Q}(\sqrt{d})$ are wonderously rich, so too are their noncommutative analogues. In this part of the text, we explore these beginnings of noncommutative algebraic number theory.

In this chapter, we begin with some prerequisites from commutative algebra, embarking on a study of integral structures and linear algebra over domains.

9.1 ▹ Integral structures

Just as we find the integers $\mathbb{Z}$ inside the rational numbers $\mathbb{Q}$, more generally we want a robust notion of integrality for possibly noncommutative algebras: this is the theory of orders over a domain.

We first have to understand the linear algebra aspects of this question. Let R be a domain with field of fractions $F := \operatorname{Frac} R$, and let V be a finite-dimensional F-vector space. An R-**lattice** in V is a finitely generated R-submodule $M \subset V$ with $MF = V$. If R is a PID (for example, $R = \mathbb{Z}$), then M is an R-lattice if and only if $M = Rx_1 \oplus \cdots \oplus Rx_n$ where $x_1, \ldots, x_n$ is a basis for V as an F-vector space.

Between M and V lies intermediate structures, where instead of allowing all denominators (in the field of fractions), we only allow certain denominators; these are the *localizations* of M. To fix ideas, suppose $R = \mathbb{Z}$, so $M \simeq \mathbb{Z}^n$; we call a $\mathbb{Z}$-lattice simply a **lattice**. For a prime p, we define the **localization** of $\mathbb{Z}$ away from p to be

$$\mathbb{Z}_{(p)} := \{a/b \in \mathbb{Q} : p \nmid b\} \subset \mathbb{Q}.$$

In the localization, we can focus on those aspects of the lattice concentrated at the prime p. Extending scalars, $M_{(p)} := M\mathbb{Z}_{(p)} \subseteq V$ is a $\mathbb{Z}_{(p)}$-lattice in V, again called the **localization** of M at p. These localizations determine the lattice M in the following strong sense (Theorem 9.4.9).

Theorem 9.1.1. (Local-global dictionary for lattices). *Let V be a finite-dimensional $\mathbb{Q}$-vector space, and let $M \subseteq V$ be a lattice. Then the map $N \mapsto (N_{(p)})_p$ establishes a*

© The Author(s) 2021

J. Voight, *Quaternion Algebras*, Graduate Texts in Mathematics 288,
https://doi.org/10.1007/978-3-030-56694-4_9

bijection between lattices N and collections of lattices $(N_{(p)})_p$ (indexed by primes p) where $M_{(p)} = N_{(p)}$ for all but finitely many primes p.

By this theorem, the choice of "reference" lattice M is arbitrary. Because of the importance of this theorem, a property of a lattice that holds if and only if it holds over every localization is called a **local property**.

Finally, often vector spaces come equipped with a measure of length, or more generally a quadratic form; we can restrict these to the lattice $M \subseteq V$ with V a $\mathbb{Q}$-vector space. More intrinsically, we define a **quadratic form** $Q\colon M \to \mathbb{Z}$ to be a map satisfying:

(i) $Q(ax) = a^2 Q(x)$ for all $a \in \mathbb{Z}$ and $x \in M$, and

(ii) The associated map $T\colon M \times M \to \mathbb{Z}$ by $T(x,y) = Q(x+y) - Q(x) - Q(y)$ is ($\mathbb{Z}$-)bilinear.

Condition (i) explains (partly) the 'quadratic' nature of the map, and part (ii) is the usual way relating norms (quadratic forms) to bilinear forms. Choosing a basis $e_1, \dots, e_n$ for $M \simeq \mathbb{Z}^n$, we may then write

$$Q(x_1 e_1 + \dots + x_n e_n) = a_{11}x_1^2 + a_{12}x_1x_2 + \dots + a_{nn}x_n^2 \in \mathbb{Z}[x_1, \dots, x_n] \qquad (9.1.2)$$

as a homogeneous polynomial of degree 2.

9.2 Bits of commutative algebra

We begin with a brief review of some bits of commutative algebra relevant to our context: we need just enough to do linear algebra over (commutative) domains with good properties. Good general references for the basic facts from algebra we use (Dedekind domains, localization, etc.) are Atiyah–Macdonald [AM69], Matsumura [Mat89, Chapter 8], Curtis–Reiner [CR81, §1, §4], Reiner [Rei2003, Chapter 1], and Bourbaki [Bou98].

Throughout this chapter, let R be a (commutative) noetherian domain with field of fractions $F := \operatorname{Frac} R$.

9.2.1. An R-module P is **projective** if it is a direct summand of a free module; equivalently, P is projective if and only if every R-module surjection $M \to P$ of R-modules has a section, i.e., an R-module homomorphism $g\colon P \to M$ such that $f \circ g = \mathrm{id}_P$.

Accordingly, a free R-module is projective. A projective R-module M is necessarily **torsion free** over R, which is to say, if $rx = 0$ with $r \in R$ and $x \in M$, then $r = 0$ or $x = 0$.

9.2.2. A **fractional ideal** of R is a nonzero finitely generated R-submodule $\mathfrak{b} \subseteq F$, or equivalently, a subset of the form $\mathfrak{b} = d\mathfrak{a}$ where $\mathfrak{a} \subseteq R$ is a nonzero ideal and $d \in F^\times$. Two fractional ideals $\mathfrak{a}, \mathfrak{b}$ of R are isomorphic (as R-modules) if and only if there exists $c \in F^\times$ such that $\mathfrak{b} = c\mathfrak{a}$: indeed, given an isomorphism $\mathfrak{a} \simeq \mathfrak{b}$, we may extend scalars to F to obtain an F-linear map $F \simeq F$, which must be given by $c \in F^\times$, and conversely.

A **Dedekind domain** is a noetherian, integrally closed domain such that every nonzero prime ideal is maximal.

Example 9.2.3. A field or a PID is a Dedekind domain; in particular, the rings $\mathbb{Z}$ and $\mathbb{F}_p[t]$ are Dedekind domains. If F is a finite extension of $\mathbb{Q}$ or $\mathbb{F}_p(t)$, then the integral closure of $\mathbb{Z}$ or $\mathbb{F}_p[t]$ in F respectively is a Dedekind domain.

9.2.4. Suppose R is a Dedekind domain. Then a finitely generated R-module is projective if and only if it is torsion free. Moreover, every nonzero ideal $\mathfrak{a}$ of R can be written uniquely as the product of prime ideals (up to reordering). For every fractional ideal $\mathfrak{a}$ of R, the set $\mathfrak{a}^{-1} := \{a \in F : a\mathfrak{a} \subseteq R\}$ is a fractional ideal with $\mathfrak{a}\mathfrak{a}^{-1} = R$. Therefore the set of fractional ideals of R forms a group under multiplication. The set of principal fractional ideals comprises a subgroup, and we define $\operatorname{Cl} R$ to be the quotient, or equivalently the group of isomorphism classes of fractional ideals of R.

9.3 Lattices

Let V be a finite-dimensional F-vector space.

Definition 9.3.1. An R-**lattice** in V is a finitely generated R-submodule $M \subseteq V$ with $MF = V$. We refer to a $\mathbb{Z}$-lattice as a **lattice**.

The condition that $MF = V$ is equivalent to the requirement that M contains a basis for V as an F-vector space.

Example 9.3.2. An R-lattice in $V = F$ is the same thing as a fractional ideal of R.

We will be primarily concerned with projective R-lattices; if R is a Dedekind domain, then a finitely generated R-submodule $M \subseteq V$ is torsion free and hence automatically projective (9.2.4).

9.3.3. If there is no ambient vector space around, we will also call a finitely generated torsion free R-module M an R-**lattice**: in this case, M is a lattice in the F-vector space $M \otimes_R F$ because the map $M \hookrightarrow M \otimes_R F$ is injective (as M is torsion free).

Remark 9.3.4. Some authors omit the second condition in the definition of an R-lattice and say that M is **full** if $MF = V$. We will not encounter R-lattices that are not full (and when we do, we call them finitely generated R-submodules), so we avoid this added nomenclature.

By definition, an R-lattice can be thought of an R-submodule that "allows bounded denominators", as follows.

Lemma 9.3.5. *Let $M \subseteq V$ be an R-lattice and let $J \subseteq V$ be a finitely generated R-submodule. Then the following statements hold.*

(a) *For all $x \in V$, there exists nonzero $r \in R$ such that $rx \in M$.*
(b) *There exists nonzero $r \in R$ such that $rJ \subseteq M$.*
(c) *J is an R-lattice if and only if there exists nonzero $r \in R$ such that $rM \subseteq J \subseteq r^{-1}M$.*

Proof. First (a). Since $FM = V$, the R-lattice M contains an F-basis $x_1, \dots, x_n$ for V, so in particular $M \supseteq Rx_1 \oplus \cdots \oplus Rx_n$. For all $x \in V$, writing x in the basis $x_1, \dots, x_n$ and clearing (finitely many) denominators, we conclude that there exists nonzero $r \in R$ such that $rx \in M$.

For (b), let $y_1, \dots, y_m$ generate J as an R-module; then for each i, there exist $r_i \in R$ nonzero such that $r_i y_i \in M$ hence $r := \prod_i r_i \neq 0$ satisfies $rJ \subseteq M$, and therefore $J \subseteq r^{-1}M$. For (c), we repeat (b) with M interchanged with J to find nonzero $s \in R$ such that $sM \subseteq J$, so then

$$rsM \subseteq sM \subseteq J \subseteq r^{-1}M \subseteq (rs)^{-1}M. \qquad \square$$

For the rest of this section, we suppose that R is a Dedekind domain and treat lattices over R; for further references, see Curtis–Reiner [CR62, §22], O'Meara [O'Me73, §81], or Fröhlich–Taylor [FT91, §II.4]. It turns out that although not every R-lattice has a basis, it can be decomposed as a direct sum, as follows.

Theorem 9.3.6. *Let R be a Dedekind domain, let $M \subseteq V$ be an R-lattice and let $y_1, \dots, y_n$ be an F-basis for V. Then there exist $x_1, \dots, x_n \in M$ and fractional ideals $\mathfrak{a}_1, \dots, \mathfrak{a}_n$ such that*

$$M = \mathfrak{a}_1 x_1 \oplus \cdots \oplus \mathfrak{a}_n x_n \tag{9.3.7}$$

and $x_j \in Fy_1 + \cdots + Fy_j$ for $j = 1, \dots, n$.

Accordingly, we say that every R-lattice M is **completely decomposable** (as a direct sum of fractional ideals), and we call the elements $x_1, \dots, x_n$ a **pseudobasis** for the lattice M with respect to the **coefficient ideals** $\mathfrak{a}_1, \dots, \mathfrak{a}_n$. The matrix with rows x_i in the basis y_i is lower triangular by construction; without loss of generality (rescaling), we may suppose that the diagonal entries are equal to 1, in which case we say that the pseudobasis for M is given in **Hermite normal form**.

More generally, if $M = \mathfrak{a}_1 x_1 + \cdots + \mathfrak{a}_m x_m$, the sum not necessarily direct, then we say that the elements x_i are a **pseudogenerating set** for M with coefficient ideals $\mathfrak{a}_i$.

Proof of Theorem 9.3.6. We argue by induction on n, the case $n = 1$ corresponding to the case of a single fractional ideal.

Let $W := Fy_1 + \cdots + Fy_{n-1}$, and let $N = M \cap W$. Then there is a commutative diagram

$$\begin{array}{ccccccccc} 0 & \longrightarrow & N & \longrightarrow & M & \longrightarrow & M/N & \longrightarrow & 0 \\ & & \downarrow & & \downarrow & & \downarrow & & \\ 0 & \longrightarrow & W & \longrightarrow & V & \longrightarrow & V/W & \longrightarrow & 0 \end{array} \tag{9.3.8}$$

Since $N = W \cap M$, we have $M/N \hookrightarrow V/W$, and $V/W \simeq F$ projecting onto Fy_n. Since M/N is nonzero and finitely generated, by 9.2.4 we conclude $M/N \simeq \mathfrak{a} \subseteq F$ is a fractional ideal, hence projective. Therefore the top exact sequence of R-modules splits (the surjection has a section), so there exists $x \in M$ such that $M = N \oplus \mathfrak{a}x$ as R-modules. The result then follows by applying the inductive hypothesis to N. $\square$

An argument generalizing that of Theorem 9.3.6 yields the following [O'Me73, 81:11].

Theorem 9.3.9. (Invariant factors). *Let R be a Dedekind domain and let $M, N \subseteq V$ be R-lattices. Then there exists a common pseudobasis $x_1, \dots, x_n$ for M, N; i.e., there exists a basis $x_1, \dots, x_n$ for V and fractional ideals $\mathfrak{a}_1, \dots, \mathfrak{a}_n$ and $\mathfrak{b}_1, \dots, \mathfrak{b}_n$ such that*

$$M = \mathfrak{a}_1 x_1 \oplus \cdots \oplus \mathfrak{a}_n x_n$$
$$N = \mathfrak{b}_1 x_1 \oplus \cdots \oplus \mathfrak{b}_n x_n$$

Moreover, letting $\mathfrak{d}_i := \mathfrak{b}_i \mathfrak{a}_i^{-1}$ we may further take $\mathfrak{d}_1 \mid \cdots \mid \mathfrak{d}_n$, and then such $\mathfrak{d}_i$ are unique.

The unique fractional ideals $\mathfrak{d}_1, \dots, \mathfrak{d}_n$ given by Theorem 9.3.9 are called the **invariant factors** of N relative to M.

9.3.10. Let $M \subseteq V$ be an R-lattice with pseudobasis as in (9.3.7). The class $[\mathfrak{a}_1 \cdots \mathfrak{a}_n] \in \operatorname{Cl} R$ is well-defined (Exercise 9.7) and called the **Steinitz class**.

In fact, if we do not require that $x_j \in Fy_1 + \cdots + Fy_i$ for $j = 1, \dots, n$ in Theorem 9.3.6, then we can find a pseudobasis for M with $\mathfrak{a}_1 = \cdots = \mathfrak{a}_{n-1} = R$, i.e.,

$$M = Rx_1 \oplus \cdots \oplus Rx_{n-1} \oplus \mathfrak{a} x_n$$

with $[\mathfrak{a}]$ the Steinitz class of M.

9.4 Localizations

Properties of a domain are governed in an important way by its localizations, and consequently the structure of lattices, orders, and algebras can often be understood by looking at their localizations (and later, completions).

For a prime ideal $\mathfrak{p} \subseteq R$, we denote by

$$R_{(\mathfrak{p})} := \{r/s \in F : s \notin \mathfrak{p}\} \subseteq F \tag{9.4.1}$$

the localization of R at $\mathfrak{p}$. (We reserve the simpler subscript notation for the completion, defined in section 9.5.)

Example 9.4.2. If $R = \mathbb{Z}$ and $\mathfrak{p} = (2)$, then $R_{(2)} = \{r/s \in \mathbb{Q} : s \text{ is odd}\}$ consists of the subring of rational numbers with odd denominator.

Since R is a domain, the map $R \hookrightarrow R_{(\mathfrak{p})}$ is an embedding and we can recover R as an intersection

$$R = \bigcap_{\mathfrak{p}} R_{(\mathfrak{p})} = \bigcap_{\mathfrak{m}} R_{(\mathfrak{m})} \subseteq F \tag{9.4.3}$$

where the intersections are over all prime ideals of R and all maximal ideals of R, respectively.

Let V be a finite-dimensional F-vector space and let $M \subseteq V$ be an R-lattice. For a prime $\mathfrak{p}$ of R, let

$$M_{(\mathfrak{p})} := MR_{(\mathfrak{p})} \subseteq V$$

be the extension of scalars of M over $R_{(\mathfrak{p})}$; identifying $V = MF \simeq M \otimes_R F$ under multiplication, we could similarly define

$$M_{(\mathfrak{p})} := M \otimes_R R_{(\mathfrak{p})}.$$

In either lens, $M_{(\mathfrak{p})}$ is an $R_{(\mathfrak{p})}$-lattice in V. In this way, M determines a collection $(M_{(\mathfrak{p})})_{\mathfrak{p}}$ indexed over the primes $\mathfrak{p}$ of R.

9.4.4. Returning to 9.2.1, a finitely generated R-module M is projective if and only if it is **locally free**, i.e., $M_{(\mathfrak{p})}$ is free for all prime ideals of R.

The ability to argue locally and then with free objects is very useful, and so very often we will restrict our attention to projective (equivalently, locally free) R-modules.

9.4.5. The localization of a Dedekind domain R is a discrete valuation ring (DVR). A DVR is equivalently a local PID that is not a field. In particular, a DVR is integrally closed and every finitely generated module over a DVR is free.

Consequently, if R is a Dedekind domain, then every fractional ideal of R is **locally principal**, i.e., if $\mathfrak{a} \subseteq F$ is a fractional ideal, then for all primes $\mathfrak{p}$ of R we have $\mathfrak{a}_{(\mathfrak{p})} = a_{\mathfrak{p}} R_{(\mathfrak{p})}$ for some $a_{\mathfrak{p}} \in F^{\times}$.

We now prove a version of the equality (9.4.3) for R-lattices (recall Definition 9.3.1).

Lemma 9.4.6. *Let M be an R-lattice in V. Then*

$$M = \bigcap_{\mathfrak{p}} M_{(\mathfrak{p})} = \bigcap_{\mathfrak{m}} M_{(\mathfrak{m})} \subseteq V$$

where the intersection is over all prime (maximal) ideals $\mathfrak{p}$.

Proof. It suffices to prove the statement for maximal ideals since $M_{(\mathfrak{m})} \subseteq M_{(\mathfrak{p})}$ whenever $\mathfrak{m} \supseteq \mathfrak{p}$. The inclusion $M \subseteq \bigcap_{\mathfrak{m}} M_{(\mathfrak{m})}$ is clear. Conversely, let $x \in V$ satisfy $x \in M_{(\mathfrak{m})}$ for all maximal ideals $\mathfrak{m}$. Let

$$(M : x) := \{r \in R : rx \in M\}.$$

Then $(M : x)$ is an ideal of R. For a maximal ideal $\mathfrak{m}$ of R, since $x \in M_{(\mathfrak{m})}$ there exists $0 \neq r_{\mathfrak{m}} \in R \smallsetminus \mathfrak{m}$ such that $r_{\mathfrak{m}} x \in M$. Thus $r_{\mathfrak{m}} \in (M : x)$ and $(M : x)$ is not contained in any maximal ideal of R. Therefore $(M : x) = R$ and hence $x \in M$. ◻

Corollary 9.4.7. *Let M, N be R-lattices in V. Then the following are equivalent:*

(i) *$M \subseteq N$;*
(ii) *$M_{(\mathfrak{p})} \subseteq N_{(\mathfrak{p})}$ for all prime ideals $\mathfrak{p}$ of R; and*
(iii) *$M_{(\mathfrak{m})} \subseteq N_{(\mathfrak{m})}$ for all maximal ideals $\mathfrak{m}$ of R.*

Proof. The implications (i) ⇒ (ii) ⇒ (iii) are direct; for the implication (iii) ⇒ (i), we have $M = \bigcap_{\mathfrak{p}} M_{(\mathfrak{m})} \subseteq \bigcap_{\mathfrak{p}} N_{(\mathfrak{m})} = N$ by Lemma 9.4.6. ◻

In particular, it follows from Corollary 9.4.7 that $M = N$ for R-lattices M, N if and only if $M_{(\mathfrak{p})} = N_{(\mathfrak{p})}$ for all primes $\mathfrak{p}$ of R.

9.4.8. A property that holds if and only if it holds locally (as in Corollary 9.4.7, for the property that one lattice is contained in another) is called a **local property**.

To conclude this section, we characterize in a simple way the conditions under which a collection $(M_{(\mathfrak{p})})_{\mathfrak{p}}$ of $R_{(\mathfrak{p})}$-lattices arise from a global R-lattice. We will see that just as a nonzero ideal of R can be factored uniquely into a product of prime ideals, and hence by the data of these primes and their exponents, so too can a lattice be understood by a finite number of localized lattices, once a "reference" lattice has been chosen (to specify the local behavior of the lattice at other primes).

Theorem 9.4.9. (Local-global dictionary for lattices). *Let R be a Dedekind domain, and let $M \subseteq V$ be an R-lattice. Then the map $N \mapsto (N_{(\mathfrak{p})})_{\mathfrak{p}}$ establishes a bijection between R-lattices $N \subseteq V$ and collections of lattices $(N_{(\mathfrak{p})})_{\mathfrak{p}}$ indexed by the primes $\mathfrak{p}$ of R satisfying $M_{(\mathfrak{p})} = N_{(\mathfrak{p})}$ for all but finitely many primes $\mathfrak{p}$.*

In Theorem 9.4.9, the choice of the "reference" lattice M is arbitrary: if M' is another lattice, then by Theorem 9.4.9 $M_{(\mathfrak{p})} = M'_{(\mathfrak{p})}$ for all but finitely many primes $\mathfrak{p}$, so we get the same set of lattices replacing M by M'. In particular, *any* lattice $N \subseteq V$ agrees with any other one at all but finitely many localizations.

Remark 9.4.10. In Theorem 9.4.9, there is a bit of notational abuse: when we write a collection $(N_{(\mathfrak{p})})_{\mathfrak{p}}$, we do not mean to imply that there is (yet) an R-lattice N such that the localization of N at $\mathfrak{p}$ is equal to $N_{(\mathfrak{p})}$. This conclusion is what is provided by the theorem (the statement of surjectivity), so the notational conflict is only temporary.

Proof of Theorem 9.4.9. Let $N \subseteq V$ be an R-lattice. Then there exists $0 \neq r \in R$ such that $rM \subseteq N \subseteq r^{-1}M$. But r is contained in only finitely many prime (maximal) ideals of R, so for all but finitely many primes $\mathfrak{p}$, the element r is a unit in $R_{(\mathfrak{p})}$ and thus $M_{(\mathfrak{p})} = N_{(\mathfrak{p})}$.

So consider the set of collections $(N_{(\mathfrak{p})})_{\mathfrak{p}}$ of lattices where $N_{(\mathfrak{p})}$ is an $R_{(\mathfrak{p})}$-lattice for each prime $\mathfrak{p}$ with the property that $M_{(\mathfrak{p})} = N_{(\mathfrak{p})}$ for all but finitely many primes $\mathfrak{p}$ of R. Given such a collection, we define $N = \bigcap_{\mathfrak{p}} N_{(\mathfrak{p})} \subseteq V$. Then N is an R-submodule of V. We show it is an R-lattice in V. For each $\mathfrak{p}$ such that $M_{(\mathfrak{p})} \neq N_{(\mathfrak{p})}$, there exists $r_{\mathfrak{p}} \in R$ such that $r_{\mathfrak{p}} M_{(\mathfrak{p})} \subseteq N_{(\mathfrak{p})} \subseteq r_{\mathfrak{p}}^{-1} M_{(\mathfrak{p})}$. Therefore, if $r = \prod_{\mathfrak{p}} r_{\mathfrak{p}}$ is the product of these elements, then $rM_{(\mathfrak{p})} \subseteq N \subseteq r^{-1}M_{(\mathfrak{p})}$ for all primes $\mathfrak{p}$ with $M_{(\mathfrak{p})} \neq N_{(\mathfrak{p})}$. On the other hand, if $M_{(\mathfrak{p})} = N_{(\mathfrak{p})}$ then already $rM_{(\mathfrak{p})} \subseteq M_{(\mathfrak{p})} = N_{(\mathfrak{p})} \subseteq r^{-1}N_{(\mathfrak{p})} = r^{-1}M_{(\mathfrak{p})}$. Therefore by Corollary 9.4.7, we have $rM \subseteq N \subseteq r^{-1}M$, and so N is an R-lattice.

By Lemma 9.4.6, the association $(N_{(\mathfrak{p})})_{\mathfrak{p}} \mapsto \bigcap_{\mathfrak{p}} N_{(\mathfrak{p})}$ is an inverse to $N \mapsto (N_{(\mathfrak{p})})_{\mathfrak{p}}$. Conversely, given a collection $(N_{(\mathfrak{p})})_{\mathfrak{p}}$, for a nonzero prime $\mathfrak{p}$, we have $\left(\bigcap_{\mathfrak{q}} N_{\mathfrak{q}}\right)_{(\mathfrak{p})} = N_{(\mathfrak{p})}$ since $(R_{\mathfrak{q}})_{(\mathfrak{p})} = F$ so $(N_{\mathfrak{q}})_{(\mathfrak{p})} = V$ whenever $\mathfrak{q} \neq \mathfrak{p}$. □

9.5 Completions

Next, we briefly define the completion and show that the local-global dictionary holds in this context as well. (We will consider completions in the context of local fields more generally starting in chapter 12, so the reader may wish to return to this section later.) For a general reference on completions (and the induced topology), see e.g.

Atiyah–Macdonald [AM69, Chapter 10], Matsumura [Mat89, Chapter 8], Bourbaki [Bou98, Chapter III, §3]. To avoid diving too deeply into commutative algebra we suppose that R is a DVR, with maximal ideal $\mathfrak{p}$: for example, we might take the localization of a Dedekind domain R at a prime ideal by 9.4.5. There is a natural system of compatible projection maps $R/\mathfrak{p}^{n+1} \to R/\mathfrak{p}^n$ indexed by integers $n \geq 1$, and we define the **completion** of R at $\mathfrak{p}$ to be the inverse (or projective) limit under this system:

$$\begin{aligned} R_{\mathfrak{p}} &:= \varprojlim_n R/\mathfrak{p}^n \\ &:= \left\{ a = (a_n)_n \in \prod_{n=1}^{\infty} R/\mathfrak{p}^n : a_{n+1} \equiv a_n \pmod{\mathfrak{p}^n} \text{ for all } n \geq 1 \right\}. \end{aligned} \tag{9.5.1}$$

The completion $R_{\mathfrak{p}}$ is again a commutative ring, and we have a natural map $R \to R_{\mathfrak{p}}$ defined by $a \mapsto (a)_n$. Since R has a discrete valuation we have $\bigcap_{n=0}^{\infty} \mathfrak{p}^n = \{0\}$, so this map is injective. Moreover, since $\mathfrak{p}$ is maximal, then in fact this inclusion factors via $R \hookrightarrow R_{(\mathfrak{p})} \hookrightarrow R_{\mathfrak{p}}$ inducing isomorphisms (Exercise 9.8)

$$R/\mathfrak{p}^e \xrightarrow{\sim} R_{(\mathfrak{p})}/\mathfrak{p}^e R_{(\mathfrak{p})} \xrightarrow{\sim} R_{\mathfrak{p}}/\mathfrak{p}^e R_{\mathfrak{p}} \tag{9.5.2}$$

for all $e \geq 1$; in particular, the operation of completion is in a sense 'stronger' than the operation of localization, and the valuation on R extends naturally to $R_{\mathfrak{p}}$, so $R = F \cap R_{\mathfrak{p}} \subseteq F_{\mathfrak{p}}$. However, once local the completion looks rather similar in the context of lattices, as follows. Let $F_{\mathfrak{p}} := F \otimes_R R_{\mathfrak{p}}$ and $V_{\mathfrak{p}} := V \otimes_F F_{\mathfrak{p}}$.

Lemma 9.5.3. *Let R be a DVR with maximal ideal $\mathfrak{p} \subseteq R$. Then the maps*

$$\begin{aligned} M_{(\mathfrak{p})} &\mapsto M_{\mathfrak{p}} = M_{(\mathfrak{p})} \otimes_{R_{(\mathfrak{p})}} R_{\mathfrak{p}} \\ M_{\mathfrak{p}} \cap V &\leftarrow\!\shortmid M_{\mathfrak{p}} \end{aligned} \tag{9.5.4}$$

are mutually inverse bijections between the set of R-lattices in V and the set of $R_{\mathfrak{p}}$-lattices in $V_{\mathfrak{p}}$.

Proof. Let $M \subseteq V$ be an R-lattice. By 9.4.5 we have $M \simeq R^n$ free over R; choose a basis $M = Rx_1 \oplus \cdots \oplus Rx_n$. Then $M_{\mathfrak{p}} = M \otimes_R R_{\mathfrak{p}} \simeq R_{\mathfrak{p}} x_1 \oplus \cdots \oplus R_{\mathfrak{p}} x_n$. Let $M' := M_{\mathfrak{p}} \cap V \subseteq V_{\mathfrak{p}}$. Then $x' \in M'$ if and only if $x' = a_1 x_1 + \cdots + a_n x_n$ with $a_i \in R_{\mathfrak{p}} \cap F = R$, so indeed $M' = M$. Conversely, let $M_{\mathfrak{p}} \subseteq V_{\mathfrak{p}}$ and let $M' := M_{\mathfrak{p}} \cap V$. Then $(M')_{\mathfrak{p}} \subseteq M_{\mathfrak{p}}$, and we prove the opposite inclusion. First, a bit of setup. Let $y_1, \dots, y_n$ be an F-basis for V, and let $N = Ry_1 \oplus \cdots \oplus Ry_n$. By Lemma 9.3.5, there exists nonzero $r \in R_{\mathfrak{p}}$ such that $rN_{\mathfrak{p}} \subseteq M_{\mathfrak{p}} \subseteq r^{-1} N_{\mathfrak{p}}$. Choosing an element $s \in R$ with the same valuation as r, we have $r/s \in R_{\mathfrak{p}}^{\times}$ so in fact may suppose that $r \in R$. Rescaling the basis vectors y_i and replacing r^2 by r we may suppose that $(rN)_{\mathfrak{p}} = rN_{\mathfrak{p}} \subseteq M_{\mathfrak{p}} \subseteq N_{\mathfrak{p}}$. From the previous paragraph, we have $rN_{\mathfrak{p}} = rN \otimes_R R_{\mathfrak{p}} \subseteq M'_{\mathfrak{p}}$. Letting $(r) = \mathfrak{p}^e$ and taking (9.5.2) on each coordinate, we have an isomorphism $\varphi\colon N/rN \simeq N_{\mathfrak{p}}/rN_{\mathfrak{p}}$ induced from the natural inclusion $N \hookrightarrow N_{\mathfrak{p}}$. Now to show the inclusion, let $y \in M_{\mathfrak{p}}$. Let $x \in N \subseteq V$ be such that $\varphi(x + rN) = y + rN_{\mathfrak{p}}$; lifting to $N_{\mathfrak{p}}$, we find that there exists $z \in rN_{\mathfrak{p}} \subseteq (M')_{\mathfrak{p}} \subseteq M_{\mathfrak{p}}$ such that $x = y + z \in M_{\mathfrak{p}} \cap V$, so $y = x - z \in (M')_{\mathfrak{p}}$. ∎

In particular, Lemma 9.5.3 implies that in the local-global dictionary for lattices over a Dedekind domain R (Theorem 9.4.9), we may also work with collections of $R_{\mathfrak{p}}$-lattices $(N_{\mathfrak{p}})_{\mathfrak{p}}$ over the completions at primes.

9.6 Index

Continuing with R a noetherian domain, let $M, N \subseteq V$ be R-lattices.

Definition 9.6.1. The R-**index** of N in M, written $[M : N]_R$, is the R-submodule of F generated by the set

$$\{\det(\delta) : \delta \in \operatorname{End}_F(V) \text{ and } \delta(M) \subseteq N\}. \tag{9.6.2}$$

The style of Definition 9.6.1, given by a large generating set (9.6.2), is the replacement for being able to work with given bases; this style will be typical for us in what follows. The determinants $\det(\delta)$ are meant in the intrinsic sense, but can be computed as the determinant of a matrix upon choosing a basis for V.

Lemma 9.6.3. *The index $[M : N]_R$ is a nonzero R-module, and if $\alpha \in \operatorname{Aut}_F(V)$ then $[\alpha M : N] = \det(\alpha)^{-1}[M : N]$.*

Proof. Exercise 9.10. □

Lemma 9.6.4. *If M, N are free (as R-submodules), then $[M : N]_R$ is a free R-module generated by the determinant of any $\delta \in \operatorname{End}_F(V)$ giving a change of basis from M to N.*

Proof. Let $x_1, \dots, x_n$ be an R-basis for M, thereby an F-basis for V. Let $y_1, \dots, y_n$ be an R-basis for N; then the map $x_i \mapsto y_i$ first extends to an R-linear isomorphism $M \xrightarrow{\sim} N$ and thereby to an F-linear map $\delta \in \operatorname{End}_F(V)$, and of course $\delta(M) \subseteq N$ by construction, so $\det(\delta) \in [M : N]_R$. Conversely, let $\delta' \in \operatorname{End}_F(V)$ be such that $\delta'(M) \subseteq N$. The map $\delta'\delta^{-1} \colon N \to N$ is an R-linear map, so $\det(\delta'\delta^{-1}) = \det(\delta')\det(\delta)^{-1} \in R$, so $\det(\delta') \in \det(\delta)R$. □

Example 9.6.5. If $N = rM$ with $r \in R$, then $[M : N]_R = r^n R$ where $n = \dim_F V$.

Example 9.6.6. If $R = \mathbb{Z}$ and $N \subseteq M$, then $[M : N]_{\mathbb{Z}}$ is the ideal generated by $\#(M/N)$, the usual index taken as abelian groups. In this case, for convenience we will often identify $[M : N]_{\mathbb{Z}}$ with its unique positive generator.

Forming the R-index commutes with localization, as follows.

Lemma 9.6.7. *Let $\mathfrak{p}$ be a prime of R. Then*

$$[M_{(\mathfrak{p})} : N_{(\mathfrak{p})}]_{R_{(\mathfrak{p})}} = ([M : N]_R)_{(\mathfrak{p})}.$$

Proof. If $\delta(M) \subseteq N$, then $\delta(M_{(\mathfrak{p})}) \subseteq N_{(\mathfrak{p})}$ by $R_{(\mathfrak{p})}$-linearity, giving the inclusion ($\supseteq$). For ($\subseteq$), let $\delta \in \operatorname{End}_F(V)$ be such that $\delta(M_{(\mathfrak{p})}) \subseteq N_{(\mathfrak{p})}$. For any $x \in M$, we have $\delta(x) \in \delta(M) \subseteq \delta(M_{(\mathfrak{p})}) \subseteq N_{(\mathfrak{p})}$, so there exists $y \in N$ and $s \in R \smallsetminus \mathfrak{p}$ such that $s\delta(x) = y \in N$. Let $x_1, \dots, x_m$ generate M as an R-module, and for each i, let $s_i \in R \smallsetminus \mathfrak{p}$ be such that $s_i\delta(x_i) \in N$. Let $s := \prod_i s_i$. Then $s\delta(M) \subseteq N$, so $\det(s\delta) = s^n \det(\delta) \in [M : N]_R$, if $n := \dim_F V$. Finally, $s \in R_{(\mathfrak{p})}^\times$, we conclude that $\det \delta \in ([M : N]_R)_{(\mathfrak{p})}$, as desired. □

Proposition 9.6.8. *Suppose that M, N are projective R-modules. Then $[M : N]_R$ is a projective R-module. Moreover, if $N \subseteq M$ then $[M : N]_R = R$ if and only if $M = N$.*

Proof. Let $\mathfrak{p}$ be a prime of R and consider the localization $([M : N]_R)_{(\mathfrak{p})}$ at $\mathfrak{p}$. Since M, N are projective R-modules, they are locally free (9.2.1). By Lemma 9.6.4, the local index $[M_{(\mathfrak{p})} : N_{(\mathfrak{p})}]_{R_{(\mathfrak{p})}}$ is a principal $R_{(\mathfrak{p})}$-ideal. By Lemma 9.6.7, we conclude that $[M : N]_R$ is locally principal, therefore projective.

The second statement follows in a similar way: we may suppose that R is local and thus $N \subseteq M$ are free, in which case $M = N$ if and only if a change of basis matrix from N to M has determinant in $R^\times$. □

For Dedekind domains, the R-index can be described as follows.

Lemma 9.6.9. *If R is a Dedekind domain and $N \subseteq M$, then $[M : N]_R$ is the product of the invariant factors (or elementary divisors) of the torsion R-module M/N.*

Proof. Exercise 9.12. □

9.7 Quadratic forms

In setting up an integral theory, we will also have need of an extension of the theory of quadratic forms integrally, generalizing those over fields (Section 4.2). For further reading on quadratic forms over rings, we suggest the books by O'Meara [O'Me73], Knus [Knu88], and Scharlau [Scha85].

Definition 9.7.1. A **quadratic map** is a map $Q \colon M \to N$ between R-modules, satisfying:

(i) $Q(rx) = r^2 Q(x)$ for all $r \in R$ and $x \in M$; and

(ii) The map $T \colon M \times M \to N$ defined by

$$T(x, y) = Q(x + y) - Q(x) - Q(y)$$

is R-bilinear.

The map T in (ii) is called the **associated bilinear map**.

Remark 9.7.2. The bilinearity condition (ii) can be given purely in terms of Q: we require

$$Q(x + y + z) = Q(x + y) + Q(x + z) + Q(y + z) - Q(x) - Q(y) - Q(z)$$

for all $x, y, z \in M$.

Definition 9.7.3. A **quadratic module** over R is a quadratic map $Q\colon M \to L$ where M is a projective R-module of finite rank and L is a projective R-module of rank 1. A **quadratic form** over R is a quadratic module with codomain $L = R$.

A quadratic module $Q\colon M \to L$ is **free** if M and L are free as R-modules, and a quadratic form $Q\colon M \to R$ is **free** if M is free as an R-module.

Example 9.7.4. Let $Q\colon V \to F$ be a quadratic form. Let $M \subseteq V$ be an R-lattice such that $Q(M) \subseteq L$ where L is an invertible R-module. (When R is a Dedekind domain, we may take $L = Q(M)$, see Exercise 9.13.) Then the restriction $Q|_M\colon M \to L$ is a quadratic module over R.

Conversely, if $Q\colon M \to L$ is a quadratic module over R, then the extension $Q\colon M \otimes_R F \to L \otimes_R F \simeq F$ is a quadratic form over F. Moreover, at the slight cost of some generality (replacing an object by an isomorphic one), by choosing an isomorphism $L \otimes_R F \simeq F$ we may suppose that Q takes values in an invertible fractional ideal $\mathfrak{l} \subseteq F$.

Example 9.7.5. If $Q\colon M \to L$ is a quadratic module and $\mathfrak{a} \subseteq R$ is a projective R-ideal, then Q extends naturally by property (i) to a quadratic module $\mathfrak{a}M \to \mathfrak{a}^2 L$.

Definition 9.7.6. A **similarity** between two quadratic modules $Q\colon M \to L$ and $Q'\colon M' \to L'$ is a pair of R-module isomorphisms $f\colon M \xrightarrow{\sim} M'$ and $h : L \xrightarrow{\sim} L'$ such that $Q'(f(x)) = h(Q(x))$ for all $x \in M$, i.e., such that the diagram

$$\begin{array}{ccc} M & \xrightarrow{Q} & L \\ \wr\downarrow f & & \wr\downarrow h \\ M' & \xrightarrow{Q'} & L' \end{array} \tag{9.7.7}$$

commutes. An **isometry** between quadratic modules is a similarity with $L = L'$ and h the identity map.

Definition 9.7.8. Let $Q\colon M \to L$ be a quadratic module over R. Then Q is **nondegenerate** if the R-linear map

$$\begin{aligned} T\colon M &\to \operatorname{Hom}_R(M, L) \\ x &\mapsto (y \mapsto T(x, y)) \end{aligned} \tag{9.7.9}$$

is injective; and Q is **nonsingular** (or **regular**) if the map (9.7.9) is an isomorphism.

Example 9.7.10. If $R = F$ is a field, then (by linear algebra) Q is nondegenerate if and only if Q is nonsingular.

Example 9.7.11. A quadratic module is nondegenerate if and only if its base extension

$$Q_F : M \otimes_R F \to L \otimes_R F \simeq F$$

is nondegenerate, since the kernel can be detected over F. Recalling the definition of discriminant (Definition 4.3.3 for char $F \neq 2$ and Definition 6.3.1 in general), we conclude that Q is nondegenerate if and only if disc $Q_F \neq 0$.

The apparent notion of discriminant of a quadratic module needs some care in its definition in this generality; it is delayed until section 15.3, where discriminantal notions are explored in some detail.

Example 9.7.12. Borrowing from the future (see Lemma 15.3.8): if $M \simeq R^n$ is free, then choosing a basis for M and computing (half-)discriminant $\operatorname{disc} Q$, we will see that M is nonsingular if and only if $\operatorname{disc} Q \in R^\times$.

We now define the notions of genus and classes.

Definition 9.7.13. Let $Q\colon M \to L$ be a quadratic module. The **genus** $\operatorname{Gen} Q$ is the set of quadratic modules that are locally isometric to Q, i.e., $Q'_{(\mathfrak{p})} \sim Q_{(\mathfrak{p})}$ for all primes $\mathfrak{p} \subseteq R$. The **class set** $\operatorname{Cl} Q$ is the set of isometry classes in the genus.

We conclude with some comments on the codomain of a quadratic map.

Definition 9.7.14. A quadratic module $Q\colon M \to L$ is **primitive** if $Q(M)$ generates L as an R-module.

9.7.15. If $Q\colon R^n \to R$ is a quadratic form, written

$$Q(x_1, \dots, x_n) = \sum_{1 \le i \le j \le n} a_{ij} x_i x_j \in R[x_1, \dots, x_n],$$

then Q is primitive if and only if the coefficients a_{ij} generate the unit ideal R.

If R is a Dedekind domain, then $Q(M) \subseteq L$ is again projective (locally at a prime generated by an element of minimal valuation), so one can always replace $Q\colon M \to L$ by $Q\colon M \to Q(M)$ to get a primitive quadratic module; when R is a PID, up to similarity we may divide through by greatest common divisor of the coefficients a_{ij} in the previous paragraph.

9.7.16. In our admittedly abstract treatment of quadratic modules so far, we have specifically allowed the *codomain* of the quadratic map to vary at the same time as the domain—in particular, we do not ask that they necessarily take values in R.

Remark 9.7.17. In certain lattice contexts with R a Dedekind domain, a quadratic form with values in a fractional ideal $\mathfrak{a}$ is called an **$\mathfrak{a}$-modular quadratic form**. Given the overloaded meanings of the word *modular*, we do not employ this terminology. In the geometric context, a quadratic module is called a *line-bundle valued quadratic form*. Whatever the terminology, we will see in Chapter 22 that it is important to keep track of the codomain of the quadratic map just as much as the domain, and in particular we cannot assume that either is free when R is not a PID.

9.8 Normalized form

To conclude this chapter, we discuss an explicit *normalized form* for quadratic forms. Let R be a local PID; then R is either a field or a DVR. In either case, R has valuation $v\colon R \to \mathbb{Z}_{\ge 0} \cup \{\infty\}$ and uniformizer π; when R is a field, we take a trivial valuation and $\pi = 1$.

Let $Q\colon M \to R$ be a quadratic form over R. Then since R is a PID, $M \simeq R^n$ is free. We compute a basis for M in which Q has a particularly nice form, diagonalizing Q as far as possible. In cases where $2 \in R^\times$, we can accomplish a full diagonalization; otherwise, we can at least break up the form as much as possible, as follows. For $a, b, c \in R$, the quadratic form $Q(x, y) = ax^2 + bxy + cy^2$ on R^2 is denoted $[a, b, c]$.

Definition 9.8.1. A quadratic form Q over R is **atomic** if either:

(i) $Q \simeq \langle a \rangle$ for some $a \in R^\times$, or

(ii) $2 \notin R^\times$ and $Q \simeq [a, b, c]$ with $a, b, c \in R$ satisfying

$$v(b) < v(2a) \leq v(2c) \quad \text{and} \quad v(a)v(b) = 0.$$

In case (ii), we necessarily have $v(2) > 0$ and $v(b^2 - 4ac) = 2v(b)$.

Example 9.8.2. Suppose $R = \mathbb{Z}_2$ is the ring of 2-adic integers, so that $v(x) = \operatorname{ord}_2(x)$ is the largest power of 2 dividing $x \in \mathbb{Z}_2$. Recall that $\mathbb{Z}_2^\times/\mathbb{Z}_2^{\times 2}$ is represented by the elements $\pm 1, \pm 5$, therefore a quadratic form Q over $\mathbb{Z}_2$ is atomic of type (i) above if and only if $Q(x) \simeq \pm x^2$ or $Q(x) \simeq \pm 5x^2$. For forms of type (ii), the conditions $v(b) < v(2a) = v(a) + 1$ and $v(a)v(b) = 0$ imply $v(b) = 0$, and so a quadratic form Q over $\mathbb{Z}_2$ is atomic of type (ii) if and only if $Q(x, y) \simeq ax^2 + xy + cy^2$ with $\operatorname{ord}_2(a) \leq \operatorname{ord}_2(c)$. Replacing x by ux and y by $u^{-1}y$ for $u \in \mathbb{Z}_2^\times$ we may suppose $a = \pm 2^t$ or $a = \pm 5 \cdot 2^t$ with $t \geq 0$, and then the atomic representative $[a, 1, c]$ of the isomorphism class of Q is unique.

A quadratic form Q is **decomposable** if Q can be written as the orthogonal sum of two quadratic forms ($Q \simeq Q_1 \boxplus Q_2$) and is **indecomposable** otherwise. It follows by induction on the rank of M that Q is the orthogonal sum of indecomposable forms. We will soon give an algorithmic proof of this fact and write each indecomposable form as a scalar multiple of an atomic form. We begin with the following lemma.

Lemma 9.8.3. *An atomic form Q is indecomposable.*

Proof. If Q is atomic of type (i) then the space underlying Q has rank 1 and is therefore indecomposable. Suppose $Q = [a, b, c]$ is atomic of type (ii) and assume for purposes of contradiction that Q is decomposable. It follows that if $x, y \in M$ then $T(x, y) \in 2R$. Thus we cannot have $v(b) = 0$, so $v(a) = 0$, and further $v(b) \geq v(2) = v(2a)$; this contradicts the fact that Q is atomic. □

Proposition 9.8.4. *Let R be a local PID and let $Q\colon M \to R$ be a quadratic form. Then there exists a basis of M such that the form Q can be written*

$$Q \simeq \pi^{e_1} Q_1 \boxplus \dots \boxplus \pi^{e_n} Q_n$$

where the forms Q_i are atomic and $0 \leq e_1 \leq \dots \leq e_n \leq \infty$.

In the above proposition, we interpret $\pi^\infty = 0$. A form as presented in Proposition 9.8.4 is called **normalized**; this normalized form need not be unique.

Proof. When $R = F$ is a field with $\operatorname{char} F \neq 2$, we are applying the standard method of Gram–Schmidt orthogonalization to diagonalize the quadratic form. This argument can be adapted to the case where $R = F$ is a field with $\operatorname{char} F = 2$, see e.g. Scharlau [Scha85, §9.4]. For the general case, we make further adaptations to this procedure: see Voight [Voi2013, Algorithm 3.12] for a constructive (algorithmic) approach. □

Exercises

Let R be a noetherian domain with field of fractions $F := \operatorname{Frac} R$.

1. Let V be a finite-dimensional F-vector space and let $M, N \subseteq V$ be R-lattices. Show that $M + N$ and $M \cap N$ are R-lattices.
▹ 2. Let B be an F-algebra and let $I \subset B$ be an R-lattice. Show that there exists a nonzero $r \in R \cap I$.
3. Give an example of a non-noetherian ring R and modules $N \subset M$ such that M is finitely generated but N is not finitely generated.
4. Let k be a field and $R = k[x, y]$. Show that the R-module (x, y) is not projective.
5. Let R be a Dedekind domain. Show that every ideal of R is projective, as follows. Let $\mathfrak{a} \subseteq R$ be a nonzero ideal. (The zero ideal is trivially projective.) Since $\mathfrak{a}\mathfrak{a}^{-1} = R$, we may write $1 = \sum_{i=1}^{n} a_i b_i$ with $a_i \in \mathfrak{a}$ and $b_i \in \mathfrak{a}^{-1}$.

   (a) Define the map $\phi\colon R^n \to \mathfrak{a}$ by $\phi(x_1, \dots, x_n) = \sum_{i=1}^{n} a_i x_i$. Observe that ϕ is an R-module homomorphism, and construct a right inverse ψ to ϕ, i.e., $\phi\psi = \mathrm{id}_{\mathfrak{a}}$.

   (b) Using (a), show that $\mathfrak{a}$ is a direct summand of R^n, so $\mathfrak{a}$ is projective.

6. Let $\mathfrak{m} \subset R$ be a maximal ideal and let M be a finitely generated R-module. Let

$$\operatorname{ann}_R M := \{r \in R : rx = 0 \text{ for all } x \in M\}$$

   be the **annihilator** of M. Show that $M_{(\mathfrak{m})} = \{0\}$ if and only if $\mathfrak{m} + \operatorname{ann}_R M = R$.
7. Suppose R is a Dedekind domain. Let V be a finite-dimensional F-vector space and let $M \subseteq V$ be an R-lattice. Given a pseudobasis $M = \mathfrak{a}_1 x_1 \oplus \cdots \oplus \mathfrak{a}_n x_n$ as in (9.3.7), let $[\mathfrak{a}_1 \cdots \mathfrak{a}_n] \in \operatorname{Cl} R$. Show that this class (the Steinitz class, 9.3.10) is well-defined for M independent of the choice of pseudobasis.
8. Let R be a DVR with maximal ideal $\mathfrak{m}$. Show that if $s \notin \mathfrak{m}$ then $1/s \in R_{\mathfrak{m}}$, so there are natural inclusions

$$R \hookrightarrow R_{(\mathfrak{m})} \hookrightarrow R_{\mathfrak{m}}$$

   from the domain into its localization into the completion, inducing isomorphisms

$$R/\mathfrak{p}^e \xrightarrow{\sim} R_{(\mathfrak{p})}/\mathfrak{p}^e R_{(\mathfrak{p})} \xrightarrow{\sim} R_{\mathfrak{p}}/\mathfrak{p}^e R_{\mathfrak{p}}$$

   for all $e \geq 1$.
▹ 9. Let V be a finite-dimensional vector space over F and $M \subseteq V$ an R-lattice. Let $\mathfrak{p}$ be a prime of R. Show that if $M_{(\mathfrak{p})} \subseteq V$ is an $R_{(\mathfrak{p})}$-lattice then $M_{\mathfrak{p}} \cap V = M_{(\mathfrak{p})}$. Conclude that Lemma 9.5.3 holds.

10. Let V be a finite-dimensional F-vector space and let $M, N \subseteq V$ be R-lattices.

    (a) Show that the index $[M : N]_R$ is a nonzero R-module. *[Hint: use Lemma 9.3.5.]*

    (b) For $\alpha \in \mathrm{Aut}_F(V)$, show $[\alpha M : N] = \det(\alpha)^{-1}[M : N]$.

11. Find R-lattices $M, N \subseteq V$ such that $[M : N]_R = R$ but $M \neq N$.
12. Prove Lemma 9.6.9, as follows. Suppose R is a Dedekind domain, and let $N \subseteq M \subseteq V$ be R-lattices in a finite-dimensional vector space V over F. Prove that $[M : N]_R$ is the product of the invariant factors (or elementary divisors) of the torsion R-module M/N.
13. Suppose R is a Dedekind domain. Let $Q \colon V \to F$ be a quadratic form over F, let $M \subseteq V$ be an R-lattice, and let $L := Q(M) \subseteq F$ be the R-submodule of F generated by the values of Q. Show that L is a fractional R-ideal.
14. Consider the ternary quadratic form $Q(x, y, z) = xy + xz$ over $\mathbb{Z}_2$. Compute a normalized form for Q.
15. Consider the following 'counterexamples' to Theorem 9.4.9 for more general integral domains as follows. Let $R = \mathbb{Q}[x, y]$ be the polynomial ring in two variables over $\mathbb{Q}$, so that $F = \mathbb{Q}(x, y)$. Let $V = F$ and $I = R$.

    (a) Show that yR has the property that $yR_{\mathfrak{p}} \neq R_{\mathfrak{p}}$ for infinitely many prime ideals $\mathfrak{p}$ of R.

    (b) Consider the collection of lattices given by $J_{\mathfrak{p}} = f(x)R_{\mathfrak{p}}$ if $\mathfrak{p} = (y, f(x))$ where $f(x) \in \mathbb{Q}[x]$ is irreducible and $J_{\mathfrak{p}} = R_{\mathfrak{p}}$ otherwise. Show that $\bigcap_{\mathfrak{p}} J_{\mathfrak{p}} = (0)$.

    [Instead, to conclude that a collection $(J_{\mathfrak{p}})_{\mathfrak{p}}$ of $R_{\mathfrak{p}}$-lattices arises from a global R-lattice J, one needs that the collection forms a *sheaf*.]
16. In this advanced exercise, we consider generalizations of the notion of lattices to a geometric context; we assume background in algebraic geometry at the level of Hartshorne [Har77, Chapter II].

    Let X be a separated, integral scheme—so for each open U, the ring $\mathscr{O}_X(U)$ is a(n integral) domain—and let $\mathscr{O}_X$ be its structure sheaf. Let F be the function field of X (so $F = \mathscr{O}_X(\{\eta\})$ where η is the generic point of X). Let V be a finite-dimensional F-vector space.

    Define a sheaf of **$\mathscr{O}_X$-lattices** in V (also called an **$\mathscr{O}_X$-lattice** in V), to be a sheaf $\mathscr{M}$ of $\mathscr{O}_X$-modules such that for each affine open set $U \subseteq X$, the set $\mathscr{M}(U)$ is an $\mathscr{O}_X(U)$-lattice in V. As usual, for $P \in X$ a point, we denote by $\mathscr{M}_{(P)}$ the stalk of $\mathscr{M}$ at P.

    (a) Show that a sheaf of $\mathscr{O}_X$-lattices in V is naturally a subsheaf of the constant sheaf V over X.
    (b) Let $X = \bigcup_i U_i$ be an affine open cover of X, with $U_i = \operatorname{Spec} R_i$. Since X is separated, each intersection $U_i \cap U_j = \operatorname{Spec} R_{ij}$ is affine, so there are natural inclusions $R_i, R_j \hookrightarrow R_{ij} \subseteq F$ of rings for each i, j. Show that a sheaf of $\mathscr{O}_X$-lattices is specified uniquely by R_i-lattices $M_i \subseteq V$ for each i,

subject to the condition that $M_i R_{ij} = M_j R_{ij}$ for each i, j. [Hint: this is an easy case of gluing, where isomorphism is replaced by equality in V.]

(c) Now suppose further that X is noetherian, normal, and of dimension ≤ 1 (also called a **Dedekind scheme**). Then the local rings of X at closed points are DVRs with fraction field F, and nonempty affine open subsets of X are the complements of finite subsets of closed points and of the form $U = \operatorname{Spec} R$ with R an Dedekind domain. (For example, we may take $X = \operatorname{Spec} R$ for R a Dedekind domain or X a smooth projective integral curve over a field.)

Extend the local-global dictionary for lattices to X, in the following way. Let $U = \operatorname{Spec} R \subseteq X$ be a nonempty affine open subset, and let $M \subseteq V$ be an R-lattice. Show that the map $\mathscr{N} \to (\mathscr{N}_{(P)})_P$ establishes a bijection between $\mathscr{O}_X$-lattices $\mathscr{N}$ in V and collections of lattices $(N_{(P)})_P$ indexed by the points $P \in X$, such that for all but finitely many $P \in U$ given by the prime $\mathfrak{p} \subseteq R$, we have $M_{(\mathfrak{p})} = N_{(P)} \subseteq V$.

This chapter is licensed under the terms of the Creative Commons Attribution-NonCommercial 4.0 International License (http://creativecommons.org/licenses/by-nc/4.0/), which permits any noncommercial use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license and indicate if changes were made.

The images or other third party material in this chapter are included in the chapter's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the chapter's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.

Chapter 10

Orders

In this chapter, continuing with a second background installment, we study when lattices over a domain are closed under a multiplication law: these will be orders, an integral analogue of algebras over fields.

10.1 ▹ Lattices with multiplication

We begin with a brief indication of the theory of orders over the integers. Let B be a finite-dimensional $\mathbb{Q}$-algebra. An **order** $O \subset B$ is a lattice that is also a subring of B (in particular, $1 \in O$). The property of being an order is a local property for a lattice, i.e., one may check that it is closed under multiplication in every localization $O_{(p)}$, for p prime.

An order is **maximal** if it is not properly contained in another order. For example, if we start with the quaternion algebra $B := \left(\dfrac{a,b}{\mathbb{Q}}\right)$ with $a, b \in \mathbb{Z}$ nonzero, then the lattice

$$O := \mathbb{Z} + \mathbb{Z}i + \mathbb{Z}j + \mathbb{Z}ij \subseteq B$$

is closed under multiplication, and so defines an order—but it is never a maximal order.

An important construction of lattices comes about as follows: if $I \subseteq B$ is a lattice, then

$$O_{\mathsf{L}}(I) := \{\alpha \in B : \alpha I \subseteq I\}$$

is an order, called the **left order** of I; we similarly define the **right order**.

If $O \subset B$ is an order and $\alpha \in O$, then α is **integral** (over $\mathbb{Z}$), satisfying a monic polynomial with integer coefficients. If B is a quaternion algebra, then $\alpha \in B$ satisfies its reduced characteristic polynomial of degree 2, and α is integral if and only if $\operatorname{trd}(\alpha), \operatorname{nrd}(\alpha) \in \mathbb{Z}$ (Corollary 10.3.6). When $B = F$ is a number field, the most important order in F is the ring of integers, the set of *all* integral elements: it is the unique maximal order.

Unfortunately, this construction does not work in the noncommutative setting: the set of all integral elements does not form an order. For one thing, if $O \subseteq B$ is a

© The Author(s) 2021

J. Voight, *Quaternion Algebras*, Graduate Texts in Mathematics 288,
https://doi.org/10.1007/978-3-030-56694-4_10

maximal order and $\alpha \in B^\times$, then $\alpha\mathcal{O}\alpha^{-1} \subseteq B$ is a maximal order and when B is noncommutative, we may have $\alpha\mathcal{O}\alpha^{-1} \neq \mathcal{O}$. But there are more serious problems, as the following example indicates.

Example 10.1.1. Let $B = \mathrm{M}_2(\mathbb{Q})$ and let $\alpha = \begin{pmatrix} 0 & 1/2 \\ 0 & 0 \end{pmatrix}$ and $\beta = \begin{pmatrix} 0 & 0 \\ 1/2 & 0 \end{pmatrix}$. Then $\alpha^2 = \beta^2 = 0$, so α, β are integral over $R = \mathbb{Z}$, but $\alpha + \beta$ and $\alpha\beta$ are not integral since $\mathrm{nrd}(\alpha + \beta) = -1/4$ and $\mathrm{trd}(\alpha\beta) = 1/4$. (Such a counterexample does not require the existence of zerodivisors: see Exercise 10.10.)

Understanding orders in quaternion algebras is a major task of this second part of the text. In the simplest case $B = \mathrm{M}_2(\mathbb{Q})$, every maximal order is conjugate (and thus isomorphic) in B to $\mathrm{M}_2(\mathbb{Z})$. The reader may wish to skip ahead to Chapter 11 to get to know the Hurwitz order before returning to study orders more generally.

10.2 Orders

Throughout, let R be a domain with field of fractions $F := \mathrm{Frac}(R)$, and let B be a finite-dimensional F-algebra. For further reference about orders (as lattices), see Reiner [Rei2003, Chapter 2] and Curtis–Reiner [CR81, §§23, 26].

Definition 10.2.1. An R-**order** $\mathcal{O} \subseteq B$ is an R-lattice that is also a subring of B.

In particular, if $\mathcal{O}$ is an R-order, then since $\mathcal{O}$ is a subring we have $1 \in \mathcal{O}$, and since $\mathcal{O}$ is an R-module we have $R \subseteq \mathcal{O}$. We will primarily be concerned with R-orders that are projective as R-modules, and call them **projective** R-orders.

10.2.2. An R-**algebra** is a ring $\mathcal{O}$ equipped with an embedding $R \hookrightarrow \mathcal{O}$ whose image lies in the center of $\mathcal{O}$. An R-order $\mathcal{O}$ has the structure of an R-algebra, and if $\mathcal{O}$ is an R-algebra that is finitely generated as an R-module, then $\mathcal{O}$ is an R-order of the F-algebra $B = \mathcal{O} \otimes_R F$.

Example 10.2.3. The matrix algebra $\mathrm{M}_n(F)$ has the R-order $\mathrm{M}_n(R)$. The subring $R[G] = \bigoplus_{g \in G} Rg$ is an R-order in the group ring $F[G]$.

Example 10.2.4. Let $a, b \in R \setminus \{0\}$ and consider the quaternion algebra $B = (a, b \mid F)$. Then $\mathcal{O} = R \oplus Ri \oplus Rj \oplus Rk$ is an R-order, because it is closed under multiplication (e.g., $ik = i(ij) = aj \in \mathcal{O}$).

Let $I \subseteq B$ be an R-lattice in the F-algebra B.

10.2.5. An important construction of orders comes as follows. Let

$$\mathcal{O}_{\mathsf{L}}(I) := \{\alpha \in B : \alpha I \subseteq I\}. \tag{10.2.6}$$

Lemma 10.2.7. *$\mathcal{O}_{\mathsf{L}}(I) \subseteq B$ is an R-order.*

Proof. Then $O_{\mathsf{L}}(I)$ is an R-submodule of B which is a ring. We show it is also an R-lattice. For all $\alpha \in B$, by Lemma 9.3.5(b), there exists nonzero $r \in R$ such that $r(\alpha I) \subseteq I$, hence $O_{\mathsf{L}}(I)F = B$. Also by this lemma, there exists nonzero $s \in R$ such that $s = s \cdot 1 \in I$; thus $O_{\mathsf{L}}(I)s \subseteq I$ so $O_{\mathsf{L}}(I) \subseteq s^{-1}I$. Since R is noetherian and $s^{-1}I$ is an R-lattice so finitely generated, we conclude that $O_{\mathsf{L}}(I)$ is finitely generated and is thus an R-lattice. □

Definition 10.2.8. The order $O_{\mathsf{L}}(I) = \{\alpha \in B : \alpha I \subseteq I\}$ in (10.2.6) is called the **left order** of I. We similarly define the **right order** of I by

$$O_{\mathsf{R}}(I) := \{\alpha \in B : I\alpha \subseteq I\}.$$

Example 10.2.9. It follows from Lemma 10.2.7 that B has an R-order: the R-span of an F-basis for B defines an R-lattice, so $O_{\mathsf{L}}(I)$ is an R-order. (This is a nice way of "clearing denominators" from a multiplication table to obtain an order.)

We can read other properties about lattices from their localizations, such as in the following lemma.

Lemma 10.2.10. *Let B be a finite-dimensional F-algebra and let $I \subseteq B$ be an R-lattice. Then the following are equivalent:*

(i) *I is an R-order;*
(ii) *$I_{(\mathfrak{p})}$ is an $R_{(\mathfrak{p})}$-order for all primes $\mathfrak{p}$ of R; and*
(iii) *$I_{(\mathfrak{m})}$ is an $R_{(\mathfrak{m})}$-order for all maximal ideals $\mathfrak{m}$ of R.*

Proof. For (i) ⇒ (ii) ⇒ (iii), if I is an R-order then $I_{(\mathfrak{p})}$ is an $R_{(\mathfrak{p})}$-order for all primes $\mathfrak{p}$, hence *a fortiori* for all maximal ideals $\mathfrak{m}$.

To conclude, we prove (iii) ⇒ (i), and suppose that $I_{(\mathfrak{m})}$ is an $R_{(\mathfrak{m})}$-order for all maximal ideals $\mathfrak{m}$. Then $\bigcap_{\mathfrak{m}} I_{(\mathfrak{m})} = I$ by Lemma 9.4.6. Thus $1 \in \bigcap_{\mathfrak{m}} I_{(\mathfrak{m})} = I$, and for all $\alpha, \beta \in I$ we have $\alpha\beta \in \bigcap_{\mathfrak{m}} I_{(\mathfrak{m})} = I$, so I is a subring of B and hence an order. □

Remark 10.2.11. The hypothesis that R is noetherian is used in Lemma 10.2.7, but it is not actually needed; the fact that $O_{\mathsf{L}}(I)$ is an order follows by a process often referred to as **noetherian reduction**. A basis of B yields a multiplication table, consisting of finitely many elements of F; moreover, we know that I is finitely generated as an R-module. Writing these generators in terms of a basis we can express these generators over the basis using finitely many elements of F. Let R_0 be the subring of R generated by these finitely elements, with field of fractions F_0, let B_0 be the F_0-algebra with the same multiplication table as B; let I_0 be the R_0-submodule generated by the generators for I written over R_0. Then $B = B_0 \otimes_{F_0} F$ and $I = I_0 \otimes_{R_0} R$. But now R_0 is a finitely generated commutative algebra over its prime ring (the subring generated by 1), so by the Hilbert basis theorem, R_0 is noetherian. The argument given then shows that I_0 is finitely generated as an R_0-module, whence I is finitely generated as an R-module.

Noetherian reduction applies to many results in this text, but non-noetherian rings are not our primary concern; we retain the noetherian hypothesis for simplicity of argument and encourage the interested reader to seek generalizations (when they are possible).

10.3 Integrality

Orders are composed of integral elements, defined as follows. If $\alpha \in B$, we denote by $R[\alpha] = \sum_d R\alpha^d$ the (commutative) R-subalgebra of B generated by α.

Definition 10.3.1. An element $\alpha \in B$ is **integral** over R if α satisfies a monic polynomial with coefficients in R.

Lemma 10.3.2. *For $\alpha \in B$, the following are equivalent:*

(i) *α is integral over R;*
(ii) *$R[\alpha]$ is a finitely generated R-module;*
(iii) *α is contained in a subring A that is finitely generated as an R-module.*

Proof. This lemma is standard; the only extra detail here is to note that in (iii) we do not need to assume that the subring A is commutative: (ii) $\Rightarrow$ (iii) is immediate taking $A = R[\alpha]$, and for the converse, if $A \subseteq B$ is a subring that is finitely generated as an R-module, then $R[\alpha] \subseteq A$ and since R is noetherian and A is finitely generated as an R-module, it follows that $R[\alpha]$ is also finitely generated as an R-module. □

Corollary 10.3.3. *If $\mathcal{O}$ is an R-order, then every $\alpha \in \mathcal{O}$ is integral over R.*

10.3.4. We say R is **integrally closed** (in F) if whenever $\alpha \in F$ is integral over R, then in fact $\alpha \in R$. Inside the field F, the set of elements integral over R (the **integral closure** of R in F) forms a ring: if α, β are integral over R then $\alpha + \beta$ and $\alpha\beta$ are integral since they lie in $R[\alpha, \beta]$ which is a finitely generated submodule of F. The integral closure of R is itself integrally closed.

Lemma 10.3.5. *Suppose that R is integrally closed. Then $\alpha \in B$ is integral over R if and only if the minimal polynomial of α over F has coefficients in R.*

Proof. Let $f(x) \in R[x]$ be a monic polynomial that α satisfies, and let $g(x) \in F[x]$ be the minimal polynomial of α. Let K be a splitting field for $g(x)$, and let $\alpha_1, \dots, \alpha_n$ be the roots of $g(x)$ in K. Since $g(x) \mid f(x)$, each such α_i is integral over R, and the set of elements in K integral over R forms a ring, so each coefficient of g is integral over R and belongs to F; but since R is integrally closed, these coefficients must belong to R and $g(x) \in R[x]$. □

Corollary 10.3.6. *If B is an F-algebra with a standard involution, and R is integrally closed, then $\alpha \in B$ is integral over R if and only if* $\operatorname{trd}(\alpha), \operatorname{nrd}(\alpha) \in R$.

We may characterize orders in separable algebras as follows.

Lemma 10.3.7. *Let $\mathcal{O} \subseteq B$ be a subring of a separable F-algebra B such that $\mathcal{O}F = B$. Then $\mathcal{O}$ is an R-order if and only if every $\alpha \in \mathcal{O}$ is integral.*

Proof. Let $\mathcal{O} \subseteq B$ be a subring of an F-algebra B such that $\mathcal{O}F = B$. Recall from Theorem 7.9.4 that a separable F-algebra is a semisimple F-algebra such that the symmetric bilinear pairing $(\alpha, \beta) \mapsto \operatorname{trd}(\alpha\beta)$ is nondegenerate.

We need to show that $\mathcal{O}$ is finitely generated. Let $\alpha_1, \dots, \alpha_n$ be an F-basis for B contained in $\mathcal{O}$. If $\beta \in \mathcal{O}$ then $\beta = \sum_i a_i \alpha_i$ with $a_i \in F$. We have $\beta\alpha_i \in \mathcal{O}$ since $\mathcal{O}$ is a ring, so $\operatorname{trd}(\beta\alpha_i) = \sum_j a_j \operatorname{trd}(\alpha_j\alpha_i)$ with $\operatorname{trd}(\alpha_j\alpha_i) \in R$. Now since B is separable, the matrix $(\operatorname{trd}(\alpha_i\alpha_j))_{i,j=1,\dots,n}$ is invertible, say $r = \det(\operatorname{trd}(\alpha_i\alpha_j))$, so we can solve these equations for a_j using Cramer's rule and we find that $a_j \in r^{-1}R$. Consequently $\mathcal{O} \subseteq r^{-1}(R\alpha_1 \oplus \dots \oplus R\alpha_n)$ is a submodule of a finitely generated module so (since R is noetherian) $\mathcal{O}$ is finitely generated. □

10.4 Maximal orders

The integral closure of R in F is the largest ring containing integral elements. Accordingly, we make the following more general definition.

Definition 10.4.1. An R-order $\mathcal{O} \subseteq B$ is **maximal** if it is not properly contained in another R-order.

If B is a commutative F-algebra and R is integrally closed in F, then the integral closure S of R in K is integrally closed and therefore S is a maximal R-order in K. However, if B is noncommutative, then the set of elements in B integral over R is no longer necessarily itself a ring, and so the theory of maximal orders is more complicated. (This may seem counterintuitive at first, but certain aspects of the noncommutative situation are quite different!) The problem in the noncommutative setting is that although $R[\alpha]$ and $R[\beta]$ may be finitely generated as R-modules for $\alpha, \beta \in B$, this need not be the case for the R-algebra generated by α and β.

10.4.2. It follows from Lemma 10.3.7 that a *separable* F-algebra B has a maximal R-order, as follows. By Lemma 10.2.7, B has an R-order $\mathcal{O}$ (since it has a lattice, taking the R-span of an F-basis), so the collection of R-orders containing $\mathcal{O}$ is nonempty. Given a chain of R-orders containing $\mathcal{O}$, by Lemma 10.3.7 the union of these orders is again an R-order. Since R is noetherian, there exists a maximal element in a chain.

For the rest of this section, we restrict attention and suppose that R is a Dedekind domain. We begin by showing that the property of being a maximal order is a local property.

Lemma 10.4.3. *An R-order $\mathcal{O} \subseteq B$ is maximal if and only if $\mathcal{O}_{(\mathfrak{p})}$ is a maximal $R_{(\mathfrak{p})}$-order for all primes $\mathfrak{p}$ of R.*

Proof. If $\mathcal{O}_{(\mathfrak{p})}$ is maximal for each prime $\mathfrak{p}$ then by Corollary 9.4.7 we see that $\mathcal{O}$ is maximal. Conversely, suppose $\mathcal{O}$ is maximal and suppose that $\mathcal{O}_{(\mathfrak{p})} \subsetneq \mathcal{O}'_{(\mathfrak{p})}$ is a proper containment of orders for some nonzero prime $\mathfrak{p}$. Then the set $\mathcal{O}' = \left(\bigcap_{\mathfrak{q}\neq\mathfrak{p}} \mathcal{O}_{\mathfrak{q}}\right) \cap \mathcal{O}'_{(\mathfrak{p})}$ is an R-order properly containing $\mathcal{O}$ by Lemma 10.2.10 and Theorem 9.4.9. □

Lemma 10.4.4. *Let $\mathcal{O} \subset B$ be an R-order. Then for all but finitely many primes $\mathfrak{p}$ of R, we have that $\mathcal{O}_{(\mathfrak{p})} = \mathcal{O} \otimes_R R_{(\mathfrak{p})}$ is maximal.*

Proof. By 10.4.2, there exists a maximal order $\mathcal{O}' \supseteq \mathcal{O}$. By the local-global principle for lattices (Theorem 9.4.9), we have $\mathcal{O}'_{\mathfrak{p}} = \mathcal{O}_{\mathfrak{p}}$ for all but finitely many primes $\mathfrak{p}$. □

The structure of (maximal) orders in quaternion algebras over domains of arithmetic interest is the subject of the second part of this text.

10.5 Orders in a matrix ring

In this section, we study orders in a matrix ring; we restore generality, and let R be a noetherian domain with $F = \operatorname{Frac} R$.

The matrix ring over F is just the endomorphism ring of a finite-dimension vector space over F, and we seek a similar description for orders as endomorphism rings of lattices (cf. 10.2.5).

Let V be an F-vector space with $\dim_F V = n$ and let $B = \operatorname{End}_F(V)$. Choosing a basis of V gives an identification $B = \operatorname{End}_F(V) \simeq \mathrm{M}_n(F)$. Given an R-lattice $M \subseteq V$, we define

$$\operatorname{End}_R(M) := \{f \in \operatorname{End}_F(V) : f(M) \subseteq M\} \subseteq B. \tag{10.5.1}$$

The left order (10.2.5) is the special case of (10.5.1) where $M = I \subseteq V = B$.

Example 10.5.2. If $V = Fx_1 \oplus \dots \oplus Fx_n$ and $M = Rx_1 \oplus \dots \oplus Rx_n$, then $\operatorname{End}_R(M) \simeq \mathrm{M}_n(R)$.

More generally, if M is **completely decomposable**, i.e. $M = \mathfrak{a}_1 x_1 \oplus \dots \oplus \mathfrak{a}_n x_n$ with each $\mathfrak{a}_i \subseteq F$ invertible fractional ideals, then we have $\operatorname{End}_R(M) \subseteq \mathrm{M}_n(F)$ the subring of matrices whose ij-entry belongs to the R-module

$$\mathfrak{a}_j \mathfrak{a}_i^{-1} \simeq \operatorname{Hom}_R(\mathfrak{a}_i, \mathfrak{a}_j) \subseteq \operatorname{Hom}_F(F, F) \simeq F$$

where the isomorphisms come from multiplication. For example, if $n = 2$ then

$$\operatorname{End}_R(M) \simeq \begin{pmatrix} R & \mathfrak{a}_2\mathfrak{a}_1^{-1} \\ \mathfrak{a}_1\mathfrak{a}_2^{-1} & R \end{pmatrix} \subseteq \mathrm{M}_2(F).$$

(Note how the cross terms are aligned correctly in the multiplication!) For example, if $M = Rx_1 + \mathfrak{a}x_2$, then $\operatorname{End}_R(M) \simeq \begin{pmatrix} R & \mathfrak{a}^{-1} \\ \mathfrak{a} & R \end{pmatrix}$.

Lemma 10.5.3. *Let M be an R-lattice of V. Then* $\operatorname{End}_R(M)$ *is an R-order in $B =$* $\operatorname{End}_F(V)$.

Proof. As in the proof of Lemma 10.2.7, we conclude that $\operatorname{End}_R(M)F = B$. Let $\alpha_1, \alpha_2, \dots, \alpha_n$ be an F-basis for V and let $N = R\alpha_1 \oplus \dots \oplus R\alpha_n$. Thus $\operatorname{End}_R(N) \simeq \mathrm{M}_n(R)$ is finitely generated as an R-module.

By Lemma 9.3.5 there exists nonzero $r \in R$ such that $rN \subseteq M \subseteq r^{-1}N$. Therefore, if $\phi \in \operatorname{End}_R(M)$, so that $\phi(M) \subseteq M$, then

$$(r^2\phi)(N) = r\phi(rN) \subseteq r\phi(M) \subseteq rM \subseteq N$$

and thus $\operatorname{End}_R(M) \subseteq r^{-2}\operatorname{End}_R(N)$; since R is noetherian, this implies that $\operatorname{End}_R(M)$ is finitely generated as an R-module and $\operatorname{End}_R(M)$ is an R-order in B. ∎

Lemma 10.5.4. *Let $\mathcal{O} \subseteq B = \operatorname{End}_F(V)$ be an R-order. Then $\mathcal{O} \subseteq \operatorname{End}_R(M)$ for some R-lattice $M \subseteq V$. In particular, if $\mathcal{O} \subseteq B$ is a maximal R-order, then $\mathcal{O} = \operatorname{End}_R(M)$ for some R-lattice M.*

Proof. Quite generally, if N is *any* R-lattice in V, then $M = \{x \in N : \mathcal{O}x \subseteq N\}$ is an R-submodule of N with $FM = V$ (as in Lemma 10.2.7), thus M is an R-lattice in V and $\mathcal{O} \subseteq \operatorname{End}_R(M)$. If further $\mathcal{O}$ is maximal, then the other containment so equality holds. $\square$

Corollary 10.5.5. *If R is a PID, then every maximal R-order $\mathcal{O} \subseteq B \simeq \mathrm{M}_n(F)$ is conjugate in B to $\mathrm{M}_n(R)$.*

Proof. The isomorphism $B \simeq \mathrm{M}_n(F)$ arises from a choice of basis $x_1, \dots, x_n$ for V; letting $N = \bigoplus_{i=1}^n Rx_i$ we have $\operatorname{End}_R(N) \simeq \mathrm{M}_n(R)$. The R-order $\mathrm{M}_n(R)$ is maximal by Exercise 10.6, since a PID is integrally closed.

By Lemma 10.5.4, we have $\mathcal{O} \subseteq \operatorname{End}_R(M)$ for some R-lattice $M \subseteq V$, so if $\mathcal{O}$ is maximal then $\mathcal{O} = \operatorname{End}_R(M)$. If R is a PID then M is free as an R-module, and we can write $M = Ry_1 \oplus \cdots \oplus Ry_n$; the change of basis matrix from x_i to y_i then realizes $\operatorname{End}_R(M)$ as a conjugate of $\operatorname{End}_R(N) \simeq \mathrm{M}_n(R)$. $\square$

Exercises

Let R be a noetherian domain with field of fractions F.

1. Let $\mathfrak{c} \subseteq R$ be an ideal. Show that
$$\begin{pmatrix} R & R \\ \mathfrak{c} & R \end{pmatrix} = \left\{ \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in \mathrm{M}_2(R) : c \in \mathfrak{c} \right\} \subseteq \mathrm{M}_2(R)$$
is an R-order in $\mathrm{M}_2(F)$. Note further that if $\mathfrak{c}$ is projective (equivalently, locally free) as an R-module, then this R-order is projective as an R-module.

2. Let B be a finite-dimensional F-algebra with a standard involution and let $\mathcal{O} \subseteq B$ be an R-order. Verify that $\operatorname{nrd}\colon \mathcal{O} \to R$ is a quadratic form over R.

3. Let $\mathcal{O}, \mathcal{O}' \subseteq B$ be R-orders in an F-algebra B.
   (a) Show that $\mathcal{O} \cap \mathcal{O}'$ is an R-order.
   (b) If $\mathcal{O} \subseteq \mathcal{O}'$, show that $\mathcal{O}'^{\times} \cap \mathcal{O} = \mathcal{O}^{\times}$.

4. Let $\mathcal{O} \subset B$ be an R-order in an F-algebra B and suppose that R is integrally closed. Show that $F \cap \mathcal{O} = R$.

5. Let $A_1, \dots, A_r$ be F-algebras and let $B = A_1 \times \cdots \times A_r$. Show that $\mathcal{O} \subseteq B$ is an R-order if and only if $\mathcal{O}$ is an R-lattice in B and $\mathcal{O} \cap A_i$ is an R-order for each i.

6. Let R be integrally closed. Show that $\mathrm{M}_n(R)$ is a maximal R-order in $\mathrm{M}_n(F)$.

7. Let $B = (K, b \mid F)$ be a quaternion algebra with $b \in R$ and let S be an R-order in K. Let $\mathcal{O} = S + Sj$. Show that $\mathcal{O}$ is an R-order in B.

8. Suppose that R is a PID, and let $\mathcal{O} \subseteq B$ be an R-order in the quaternion algebra B. Let $\alpha \in \mathcal{O}$ be such that $S = R[\alpha]$ is (commutative) domain that is a maximal R-order in its field of fractions.

   (a) Show that $1, \alpha$ extends to an R-basis for $\mathcal{O}$.
   (b) If moreover S is a PID, show that there exists $\beta \in \mathcal{O}$ such that $1, \alpha, \beta, \alpha\beta$ is an R-basis for $\mathcal{O}$.

9. Let B be an F-algebra with a standard involution and let $\alpha \in B$. Show that if α is integral over R then $\operatorname{trd}(\alpha^n) \in R$ for all $n \in \mathbb{Z}_{\geq 0}$. Is the converse true?

10. Generalize Example 10.1.1: Exhibit a division quaternion algebra B over $\mathbb{Q}$ and elements $\alpha, \beta \in B$ such that α, β are integral over $\mathbb{Z}$ but both $\alpha + \beta$ and $\alpha\beta$ are not.

11. Let $\alpha \in \mathrm{M}_n(F)$ have characteristic polynomial with coefficients in R. Show that α is conjugate by an element $\beta \in \mathrm{GL}_n(F)$ to an element of $\mathrm{M}_n(R)$. Explicitly, how do you find such a matrix β?

▹ 12. Let $B = \mathrm{M}_n(F)$ and let $I \subseteq B$ be an R-lattice. Let $I^{\mathsf{t}} = \{\alpha^{\mathsf{t}} : \alpha \in I\}$ be the transpose lattice. Show that $\mathcal{O}_{\mathsf{L}}(I^{\mathsf{t}}) = \mathcal{O}_{\mathsf{R}}(I)$.

▹ 13. Let $I, J \subseteq B$ be R-lattices. Let IJ be the R-submodule of B generated by products $\alpha\beta$ where $\alpha \in I$, $\beta \in J$; i.e.,

$$IJ := \left\{\textstyle\sum_{i=1}^{k} \alpha_i\beta_i : \alpha_i \in I, \beta_i \in J\right\}.$$

   (a) Show that IJ is an R-lattice.

   (b) Let $\mathfrak{p}$ be a prime of R. Show that *products commute with localization* in the sense that

$$(IJ) \otimes_R R_{(\mathfrak{p})} = (I \otimes_R R_{(\mathfrak{p})})(J \otimes_R R_{(\mathfrak{p})}) \subseteq B_{(\mathfrak{p})} = B.$$

▹ 14. Let $\mathcal{O} \subseteq B$ be an R-order in an F-algebra B.

   (a) Show that $\mathcal{O}_{\mathsf{L}}(\mathcal{O}) = \mathcal{O}_{\mathsf{R}}(\mathcal{O}) = \mathcal{O}$.

   (b) Let $\alpha \in B^\times$, and let $\alpha\mathcal{O} = \{\alpha\beta : \beta \in \mathcal{O}\}$. Show that $\alpha\mathcal{O}$ is an R-lattice and that $\mathcal{O}_{\mathsf{L}}(\alpha\mathcal{O}) = \alpha\mathcal{O}\alpha^{-1}$.

15. Let $\mathcal{O} \subseteq B$ be an R-order in an F-algebra B. Let $\gamma \in \mathcal{O}$ and let $N : B^\times \to F^\times$ be a multiplicative map. Show that $\gamma \in \mathcal{O}^\times$ if and only if $N(\gamma) \in R^\times$, and in particular, if B has a standard involution, then $\gamma \in \mathcal{O}^\times$ if and only if $\operatorname{nrd}(\gamma) \in R^\times$.

This chapter is licensed under the terms of the Creative Commons Attribution-NonCommercial 4.0 International License (http://creativecommons.org/licenses/by-nc/4.0/), which permits any noncommercial use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license and indicate if changes were made.

The images or other third party material in this chapter are included in the chapter's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the chapter's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.

Chapter 11

The Hurwitz order

With the preceding chapters on lattices and orders in hand, we are now prepared to embark on a general treatment of quaternion algebras over number fields and the arithmetic of their orders. Before we do so, for motivation and pure enjoyment, in this chapter we consider the special case of the Hurwitz order. Not only is this appropriate in a historical spirit, it is also instructive for what follows; moreover, the Hurwitz order has certain exceptional symmetries that make it worthy of specific investigation.

11.1 ▹ The Hurwitz order

Hurwitz developed the theory of integral quaternions in a treatise [Hur19] in 1919. A more modern treasure trove of detail about quaternion groups and the Hurwitz order (as well as many other things) can be found in the book by Conway–Smith [CSm2003]; the review by Baez [Bae2005] also provides an accessible overview.

We consider in this chapter the restriction of the Hamiltonians from $\mathbb{R}$ to $\mathbb{Q}$, namely, the quaternion algebra $B = \left(\dfrac{-1,-1}{\mathbb{Q}}\right)$. We further restrict to those elements with integer coordinates

$$\mathbb{Z}\langle i, j\rangle = \mathbb{Z} + \mathbb{Z}i + \mathbb{Z}j + \mathbb{Z}k, \tag{11.1.1}$$

where $k := ij$. By Example 10.2.4, this is an order in B, called the **Lipschitz order**. In the rest of this chapter, we will work over $\mathbb{Z}$ and so we will simply refer to *lattices* and *orders*.

The Lipschitz order is not a maximal order, and maximal orders have better properties. This is analogous to the fact that the ring $\mathbb{Z}[\sqrt{-3}]$ is an order in $\mathbb{Q}(\sqrt{-3})$ but is not maximal (not integrally closed), properly contained in the better-behaved maximal order $\mathbb{Z}[(-1+\sqrt{-3})/2]$ of Eisenstein integers. The comparison with the Eisenstein integers is more than incidental: the element $\alpha = i + j + k$ satisfies $\alpha^2 + 3 = 0$, so it is natural to consider

$$\omega := \frac{-1+i+j+k}{2}$$

which satisfies $\omega^2+\omega+1 = 0$. We can enlarge the Lipschitz order to include ω—indeed, this is the only possibility.

© The Author(s) 2021

J. Voight, *Quaternion Algebras*, Graduate Texts in Mathematics 288,
https://doi.org/10.1007/978-3-030-56694-4_11

Lemma 11.1.2. *The lattice*

$$O = \mathbb{Z} + \mathbb{Z}i + \mathbb{Z}j + \mathbb{Z}\omega = \mathbb{Z}\langle i, j\rangle + \mathbb{Z}\langle i, j\rangle\omega \tag{11.1.3}$$

in B is the unique order that properly contains $\mathbb{Z}\langle i, j\rangle$, and O is maximal.

The order O in (11.1.3) is called the **Hurwitz order**, and it contains $\mathbb{Z}\langle i, j\rangle$ with index 2. Note that if $\alpha \in O$, then $\alpha \in \mathbb{Z}\langle i, j\rangle$ if and only if $\operatorname{trd}(\alpha) \in 2\mathbb{Z}$.

Proof. By Exercise 11.1, the lattice O is an order.

Suppose that $O' \supsetneq \mathbb{Z}\langle i, j\rangle$ and let $\alpha = t + xi + yj + zk \in O'$ with $t, x, y, z \in \mathbb{Q}$. Then $\operatorname{trd}(\alpha) = 2t \in \mathbb{Z}$, so by Corollary 10.3.6 we have $t \in \frac{1}{2}\mathbb{Z}$. Similarly, $\alpha i \in O'$ therefore $\operatorname{trd}(\alpha i) = -2x \in \mathbb{Z}$ and $x \in \frac{1}{2}\mathbb{Z}$, and in the same way $y, z \in \frac{1}{2}\mathbb{Z}$. Finally, $\operatorname{nrd}(\alpha) = t^2 + x^2 + y^2 + z^2 \in \mathbb{Z}$, and considerations modulo 4 imply that t, x, y, z either all belong to $\mathbb{Z}$ or to $\frac{1}{2} + \mathbb{Z}$; thus $\alpha \in O$ and so $O' = O$. □

11.1.4. We can recast this calculation in terms of the local-global dictionary for lattices (Theorem 9.1.1). Since $O[\frac{1}{2}] = \mathbb{Z}\langle i, j\rangle[\frac{1}{2}]$, for every *odd* prime p we have $O_{(p)} = \mathbb{Z}\langle i, j\rangle_{(p)}$, and $O_{(2)} \supsetneq \mathbb{Z}\langle i, j\rangle_{(2)}$.

11.2 ▹ Hurwitz units

We now consider unit groups; in this section, we take $k := ij$. An element $\gamma = t + xi + yj + zk \in \mathbb{Z}\langle i, j\rangle$ is a unit if and only if $\operatorname{nrd}(\gamma) = t^2 + x^2 + y^2 + z^2 \in \mathbb{Z}^\times$, i.e. $\operatorname{nrd}(\gamma) = 1$, and since $t, x, y, z \in \mathbb{Z}$ we immediately have

$$\mathbb{Z}\langle i, j\rangle^\times = \{\pm 1, \pm i, \pm j, \pm k\} \simeq Q_8$$

is the **quaternion group** of order 8. In a similar way, taking $\gamma \in O$ in the Hurwitz order and allowing $t, x, y, z \in \frac{1}{2}\mathbb{Z}$ so that $2t, 2x, 2y, 2z$ all have the same parity, we find that

$$O^\times = Q_8 \cup (\pm 1 \pm i \pm j \pm k)/2$$

is a group of order 24.

We have $O^\times \not\simeq S_4$ (the symmetric group on 4 letters) because there is no embedding $Q_8 \hookrightarrow S_4$. (The permutation representation $Q_8 \to S_4$ obtained by the action on the cosets of the unique subgroup $\langle -1\rangle$ of index 4 factors through the quotient $Q_8 \to Q_8/\{\pm 1\} \simeq V_4 \hookrightarrow S_4$, where V_4 is the Klein 4-group.) There are 15 groups of order 24 up to isomorphism! We identify the right one as follows.

Lemma 11.2.1. *We have $O^\times \simeq \mathrm{SL}_2(\mathbb{F}_3)$.*

Proof. We reduce modulo 3. There is a ring homomorphism

$$O \to O/3O \simeq \mathbb{F}_3\langle i, j\rangle \simeq \left(\frac{-1, -1}{\mathbb{F}_3}\right).$$

Any quaternion algebra over a finite field is isomorphic to the matrix ring by Wedderburn's little theorem (Exercises 3.16, 6.16, and 7.30). Specifically, the element

$\epsilon = i + j + k$ has $\epsilon^2 = 0 \in \mathcal{O}/3\mathcal{O}$. The left ideal I generated by ϵ is an $\mathbb{F}_3$-vector space, and we compute that it has basis ϵ and $i\epsilon = -1 - j + k$. As in (7.6.3) (Proposition 7.6.2) this yields an isomorphism

$$\begin{aligned} \mathcal{O}/3\mathcal{O} &\to \mathrm{M}_2(\mathbb{F}_3) \\ i, j &\mapsto \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}, \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix} \end{aligned}$$

(a statement that can be explicitly and independently verified in Exercise 11.3). We obtain a group homomorphism $\mathcal{O}^\times \to \mathrm{SL}_2(\mathbb{F}_3)$, since the reduced norm corresponds to the determinant and $\mathrm{nrd}(\mathcal{O}^\times) = \{1\}$, and this homomorphism is injective because if $\gamma \in \mathcal{O}^\times$ has $\gamma - 1 \in 3\mathcal{O}$ then $\gamma = 1$, by inspection. Since $\#\mathcal{O}^\times = \#\,\mathrm{SL}_2(\mathbb{F}_3) = 24$, the map $\mathcal{O}^\times \hookrightarrow \mathrm{SL}_2(\mathbb{F}_3)$ is an isomorphism. □

11.2.2. The group $\mathrm{SL}_2(\mathbb{F}_3)$ acts on the left on the set of nonzero column vectors $\mathbb{F}_3^2$ up to sign, a set of cardinality $(9-1)/2 = 4$. (More generally, $\mathrm{SL}_2(\mathbb{F}_p)$ acts on $\mathbb{P}^1(\mathbb{F}_p) := (\mathbb{F}_p^2 \smallsetminus \{(0,0)\})/\mathbb{F}_p^\times$, a set of cardinality $p+1$.) This action yields a permutation representation $\mathrm{SL}_2(\mathbb{F}_3) \to S_4$; the kernel of this map is the subgroup generated by the scalar matrix -1 and so the representation gives an injective group homomorphism from $\mathrm{PSL}_2(\mathbb{F}_3) := \mathrm{SL}_2(\mathbb{F}_3)/\{\pm 1\}$ into S_4. Since $A_4 \leq S_4$ is the unique subgroup of size $24/2 = 12$, we must have $\mathrm{PSL}_2(\mathbb{F}_3) \simeq A_4$, giving an exact sequence

$$1 \to \{\pm 1\} \to \mathcal{O}^\times \to A_4 \to 1. \tag{11.2.3}$$

11.2.4. We can also visualize the group $\mathcal{O}^\times$ and the exact sequence (11.2.3), thinking of the Hamiltonians as acting by rotations (section 2.4). Recall there is an exact sequence (Corollary 2.4.21)

$$1 \to \{\pm 1\} \to \mathbb{H}^1 \to \mathrm{SO}(3) \to 1 \tag{11.2.5}$$

obtained by the left action $\alpha \mapsto \alpha v \alpha^{-1}$ for $\alpha \in \mathbb{H}^1$ and $v \in \mathbb{H}^0 \simeq \mathbb{R}^3$; specifically, by Proposition 2.4.18, a quaternion $\alpha = \cos\theta + I(\alpha)\sin\theta$ acts by rotation through the angle 2θ about the axis $I(\alpha)$.

We have been considering

$$\mathcal{O} \hookrightarrow B = \left(\frac{-1,-1}{\mathbb{Q}}\right) \hookrightarrow B \otimes_{\mathbb{Q}} \mathbb{R} = \left(\frac{-1,-1}{\mathbb{R}}\right) = \mathbb{H}, \tag{11.2.6}$$

and we now consider the corresponding embedding of groups $\mathcal{O}^1 = \mathcal{O}^\times \hookrightarrow \mathbb{H}^1$. We are led to think of the group $\mathcal{O}^\times/\{\pm 1\} \simeq A_4$ as the group of symmetries (rigid motions) of a tetrahedron (or rather, a tetrahedron and its dual), as in Figure 11.2.7.

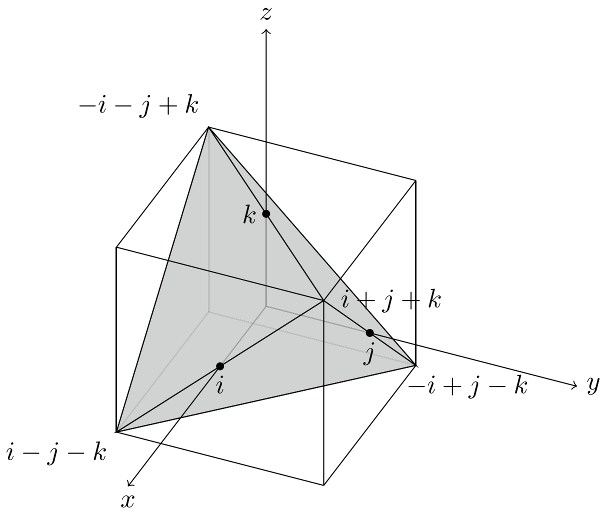


Figure 11.2.7: Symmetries of a tetrahedron, viewed quaternionically

Inside the cube in $\mathbb{R}^3$ with vertices $(\pm 1, \pm 1, \pm 1) = \pm i \pm j \pm k$, we can find four inscribed tetrahedra, for example, the tetrahedron T with vertices

$$i + j + k, i - j - k, -i + j - k, -i - j + k.$$

Then the elements $\pm i, \pm j, \pm k$ act by rotation about the x, y, z axes by an angle π (so interchanging points with the same x, y, z coordinate). The element $\pm\omega = \pm(-1 + i + j + k)/2$ rotates by the angle $2\pi/3$ fixing the point $(1, 1, 1)$ and cyclically permuting the other three points, and by symmetry we understand the action of the other elements of $\mathcal{O}^\times$. We therefore call $\mathcal{O}^\times$ the **binary tetrahedral group**. Following Conway–Smith [CSm2003, §3.3], we also write $2T = \mathcal{O}^\times$ for this group; the notation $\widetilde{A}_4$ is also used.

The subgroup $Q_8 \trianglelefteq 2T$ is normal (as it is characteristic, consisting of all elements of $\mathcal{O}$ of order dividing 4), and so we can write $2T = Q_8 \rtimes \langle\omega\rangle$ where $\langle\omega\rangle \simeq \mathbb{Z}/3\mathbb{Z}$ acts on Q_8 by conjugation, cyclically rotating the elements i, j, k. Finally, the group $2T$ has a presentation (Exercise 11.6)

$$2T \simeq \langle r, s, t \mid r^2 = s^3 = t^3 = rst = -1\rangle \tag{11.2.8}$$

via $r = i$, $s = \omega = (-1 + i + j + k)/2$, and $t = (-1 + i + j - k)/2$.

We conclude by noting that the difference between the Lipschitz and Hurwitz orders is "covered" by the extra units.

Lemma 11.2.9. *For every $\beta \in \mathcal{O}$, there exists $\gamma \in \mathcal{O}^\times$ such that $\beta\gamma \in \mathbb{Z}\langle i, j\rangle$.*

Proof. If $\beta \in \mathbb{Z}\langle i, j\rangle$ already, then we are done. Otherwise, $2\beta = t + xi + yj + zk$ with all $t, x, y, z \in \mathbb{Z}$ odd. Choosing matching signs, there exists $\gamma \in O^\times$ such that $2\beta \equiv 2\gamma \pmod{4O}$. Thus

$$(2\beta)\gamma^{-1} \equiv 2 \pmod{4O}$$

so $\beta\gamma^{-1} \in \mathbb{Z} + 2O = \mathbb{Z}\langle i, j\rangle$, so we may take γ^{-1} for the statement of the lemma. □

11.3 ▹ Euclidean algorithm

The Eisenstein order $\mathbb{Z}[(-1+\sqrt{-3})/2]$ has several nice properties. Perhaps nicest of all is that it is a Euclidean domain, so in particular it is a PID and UFD. (Alas, the ring $\mathbb{Z}[\sqrt{-3}]$ just fails to be Euclidean.)

11.3.1. The Hurwitz order also has a left (or right) Euclidean algorithm generalizing the commutative case, as follows. There is an embedding $B \hookrightarrow B \otimes_{\mathbb{Q}} \mathbb{R} \simeq \mathbb{H}$, and inside $\mathbb{H} \simeq \mathbb{R}^4$ the Hurwitz order sits as a ($\mathbb{Z}$-)lattice equipped with the Euclidean inner product, so we can think of the reduced norm by instead thinking of distance. In the Lipschitz order, we see by rounding coordinates that for all $\gamma \in B$ there exists $\mu \in \mathbb{Z}\langle i, j\rangle$ such that $\operatorname{nrd}(\gamma - \mu) \le 4 \cdot (1/2)^2 = 1$—a farthest point occurs at the center $(1/2, 1/2, 1/2, 1/2)$ of a unit cube. But this is precisely the point where the Hurwitz quaternions occur, and it follows that for all $\gamma \in B$, there exists $\mu \in O$ such that $\operatorname{nrd}(\gamma - \mu) < 1$. (In fact, we can take $\operatorname{nrd}(\gamma - \mu) \le 1/2$; see Exercise 11.7.)

Paragraph 11.3.1 becomes a right Euclidean algorithm as in the commutative case.

Lemma 11.3.2. (Hurwitz order is right norm Euclidean). *For all $\alpha, \beta \in O$ with $\beta \neq 0$, there exists $\mu, \rho \in O$ such that*

$$\alpha = \beta\mu + \rho \tag{11.3.3}$$

and $\operatorname{nrd}(\rho) < \operatorname{nrd}(\beta)$.

Proof. If $\operatorname{nrd}(\alpha) < \operatorname{nrd}(\beta)$, we may take $\mu = 0$ and $\rho = \alpha$, so suppose $\operatorname{nrd}(\alpha) \ge \operatorname{nrd}(\beta) > 0$. Let $\gamma = \beta^{-1}\alpha \in B$. Then by 11.3.1, there exists $\mu \in O$ such that $\operatorname{nrd}(\gamma - \mu) < 1$. Let $\rho = \alpha - \beta\mu$. Then by multiplicativity of the norm,

$$\operatorname{nrd}(\rho) = \operatorname{nrd}(\alpha - \beta\mu) < \operatorname{nrd}(\beta).$$

□

A similar statement to Lemma 11.3.2 holds for division on the left, i.e., in (11.3.3) we may take $\alpha = \mu\beta + \rho$ (with possibly different elements $\mu, \rho \in O$, of course).

Proposition 11.3.4. *Every right ideal $I \subseteq O$ is right principal, i.e., there exists $\beta \in I$ such that $I = \beta O$.*

Proof. Let $I \subseteq O$ be a right ideal. If $I = \{0\}$, we are done. Otherwise, there exists an element $0 \neq \beta \in I$ with minimal reduced norm $\operatorname{nrd}(\beta) \in \mathbb{Z}_{>0}$. We claim that $I = \beta O$. For all $\alpha \in I$, by the left Euclidean algorithm in Lemma 11.3.2, there exists $\mu \in O$ such that $\alpha = \mu\beta + \rho$ with $\operatorname{nrd}(\rho) < \operatorname{nrd}(\beta)$; but $\rho = \alpha - \beta\mu \in I$, so by minimality, $\operatorname{nrd}(\rho) = 0$ and $\rho = 0$, hence $\alpha = \beta\mu \in \beta O$ as claimed. □

Definition 11.3.5. Let $\alpha, \beta \in \mathcal{O}$. We say β **right divides** α (or α is a **right multiple** of β) and write $\beta \mid_{\mathsf{R}} \alpha$ if there exists $\gamma \in \mathcal{O}$ such that $\alpha = \beta\gamma$.

A **right common divisor** of $\alpha, \beta \in \mathcal{O}$ is an element $\gamma \in \mathcal{O}$ such that $\gamma \mid_{\mathsf{R}} \alpha, \beta$. A **right greatest common divisor** of α, β is a common divisor γ such that $\delta \mid_{\mathsf{R}} \gamma$ for all common divisors δ of α, β.

It follows from Lemma 11.3.2 in the same way as in the commutative case that if α, β are not both zero, then there exists a greatest common divisor of α, β, taking the last nonzero remainder in the right Euclidean algorithm.

Corollary 11.3.6. (Bézout's theorem). *For all $\alpha, \beta \in \mathcal{O}$ not both zero, there exists $\mu, \nu \in \mathcal{O}$ such that $\alpha\mu + \beta\nu = \gamma$ where γ is a right greatest common divisor of α, β.*

Proof. By Proposition 11.3.4, we may write $\alpha\mathcal{O} + \beta\mathcal{O} = \gamma\mathcal{O}$ for some $\gamma \in \mathcal{O}$, and then $\gamma \in \alpha\mathcal{O} + \beta\mathcal{O}$ implies there exists $\mu, \nu \in \mathcal{O}$ such that $\alpha\mu + \beta\nu = \gamma$. $\square$

Proposition 11.3.7. *Let $\mathcal{O}' \subset B$ be a maximal order. Then there exists $\alpha \in B^\times$ such that $\mathcal{O}' = \alpha^{-1}\mathcal{O}\alpha$, and in particular $\mathcal{O}' \simeq \mathcal{O}$ as rings.*

Proof. By clearing denominators, there exists nonzero $a \in \mathbb{Z}$ such that $a\mathcal{O}' \subseteq \mathcal{O}$. Let $I = a\mathcal{O}'\mathcal{O}$ be the right ideal of $\mathcal{O}$ generated by $a\mathcal{O}'$. Then $\mathcal{O}' \subseteq \mathcal{O}_{\mathsf{L}}(I)$, and equality holds since $\mathcal{O}'$ is maximal. By Proposition 11.3.4, we have $I = \beta\mathcal{O}$ for some $\beta \in B^\times$. We have $\mathcal{O}_{\mathsf{L}}(I) = \beta\mathcal{O}\beta^{-1}$ by Exercise 10.14, so $\mathcal{O}' = \beta\mathcal{O}\beta^{-1}$ and we may take $\alpha = \beta^{-1}$. $\square$

Example 11.3.8. The Lipschitz order $\mathbb{Z}\langle i, j\rangle$ does not enjoy the property that every right ideal is principal, as the following example shows.

Let $I = 2\mathcal{O} = 2\mathbb{Z} + 2i\mathbb{Z} + 2j\mathbb{Z} + (1 + i + j + k)\mathbb{Z}$. Then $I \subseteq \mathbb{Z}\langle i, j\rangle$ and I has the structure of a right $\mathbb{Z}\langle i, j\rangle$-ideal, in fact $I = 2\mathbb{Z}\langle i, j\rangle + (1 + i + j + k)\mathbb{Z}\langle i, j\rangle$. We claim that I is *not* principal as a right $\mathbb{Z}\langle i, j\rangle$-ideal. Indeed, suppose $I = \alpha\mathbb{Z}\langle i, j\rangle$ with $\alpha \in I$. Since $\alpha \in 2\mathcal{O}$, we have $4 \mid \operatorname{nrd}(\alpha)$. But $2 \in I$ so $2 = \alpha\beta$ with $\beta \in \mathbb{Z}\langle i, j\rangle$, so $4 = \operatorname{nrd}(2) = \operatorname{nrd}(\alpha)\operatorname{nrd}(\beta)$, whence $\operatorname{nrd}(\alpha) = 4$ and $\operatorname{nrd}(\beta) = 1$ so $\beta \in \mathbb{Z}\langle i, j\rangle^\times$ and so $2\mathcal{O} = I = \alpha\mathbb{Z}\langle i, j\rangle = 2\mathbb{Z}\langle i, j\rangle$. Cancelling the factor 2, we conclude $\mathcal{O} = \mathbb{Z}\langle i, j\rangle$, a contradiction.

For more, see Exercise 11.10.

11.4 ▹ Unique factorization

It does not follow that there is unique factorization in $\mathcal{O}$ in the traditional sense, as the order of multiplication matters. Nevertheless, there is a theory of prime factorization in $\mathcal{O}$ as follows.

Lemma 11.4.1. *Let p be prime. Then there exists $\pi \in \mathcal{O}$ such that $\pi\overline{\pi} = \operatorname{nrd}(\pi) = p$.*

Proof. We have $\operatorname{nrd}(1 + i) = 1^2 + 1^2 = 2$, so we may suppose $p \geq 3$ is odd. Then $\mathcal{O}/p\mathcal{O} \simeq (-1, -1 \mid \mathbb{F}_p) \simeq \mathrm{M}_2(\mathbb{F}_p)$ by Wedderburn's little theorem, and there exists a right ideal $I \bmod p \subset \mathcal{O}/p\mathcal{O}$ with $\dim_{\mathbb{F}_p}(I \bmod p) = 2$. Let

$$I = \{\alpha \in \mathcal{O} : \alpha \bmod p \in I \bmod p\}$$

be the preimage of I mod p in the map $\mathcal{O} \to \mathcal{O}/p\mathcal{O}$. Then $p\mathcal{O} \subsetneq I \subsetneq \mathcal{O}$. Then $I \subset \mathcal{O}$ is a right ideal, and $I \neq \mathcal{O}$. But $I = \beta\mathcal{O}$ is right principal by Proposition 11.3.4.

We claim that $\operatorname{nrd}(\beta) = p$. Since $p \in I$, we have $p = \beta\mu$ for some $\mu \in \mathcal{O}$, whence $\operatorname{nrd}(p) = p^2 = \operatorname{nrd}(\beta)\operatorname{nrd}(\mu)$ so $\operatorname{nrd}(\beta) \mid p^2$. We cannot have $\operatorname{nrd}(\beta) = 1$ or $\operatorname{nrd}(\beta) = p^2$, as these would imply $I = \mathcal{O}$ or $I = p\mathcal{O}$, impossible. We conclude that $\operatorname{nrd}(\beta) = p$. □

Remark 11.4.2. Once we have developed a suitable theory of norms, the proof that $\operatorname{nrd}(\beta) = p$ above will be immediate: if we define $\mathsf{N}(I) := \#(\mathcal{O}/I)$ then $\mathsf{N}(I) = p^2$ by construction, and it turns out that $\mathsf{N}(I) = \operatorname{nrd}(\beta)^2$.

Theorem 11.4.3. (Lagrange). *Every integer $n \geq 0$ is the sum of four squares, i.e., there exist $t, x, y, z \in \mathbb{Z}$ such that $n = t^2 + x^2 + y^2 + z^2$.*

Proof. We seek an element $\beta \in \mathbb{Z}\langle i, j\rangle$ such that $\operatorname{nrd}(\beta) = n$. By multiplicativity of the reduced norm, it is sufficient to treat the case where $n = p$ is prime. We obtain $\pi \in \mathcal{O}$ such that $\operatorname{nrd}(\pi) = p$ by Lemma 11.4.1. But now the result follows from Lemma 11.2.9, as there exists $\gamma \in \mathcal{O}^\times$ such that $\pi\gamma \in \mathbb{Z}\langle i, j\rangle$. □

Remark 11.4.4. A counterpart to Lagrange's theorem (Theorem 11.4.3) is the following theorem of Legendre and Gauss on sums of three squares: Every integer n that is not of the form $n = 4^a m$ with $m \equiv 7 \pmod 8$ can be written as the sum of three squares $n = x^2 + y^2 + z^2$. We will revisit this classical theorem in Chapter 30 as motivation for the study of *embedding numbers*, and the number of such representations will be given in terms of class numbers, following Gauss. A direct proof of the three square theorem is given by Mordell [Mor69, §20, Theorem 1], but he notes that "no really elementary treatment [of this theorem] is known".

We finish this section with a discussion of 'unique factorization' in the Hurwitz order.

Definition 11.4.5. An element $\pi \in \mathcal{O}$ is **irreducible** if whenever $\pi = \alpha\beta$ with $\alpha, \beta \in \mathcal{O}$ then either $\alpha \in \mathcal{O}^\times$ or $\beta \in \mathcal{O}^\times$.

Lemma 11.4.6. *Let $\pi \in \mathcal{O}$. Then π is irreducible if and only if $\operatorname{nrd}(\pi) = p \in \mathbb{Z}$ is prime.*

Proof. If $\operatorname{nrd}(\pi) = p$ is prime and $\pi = \alpha\beta$ then $\operatorname{nrd}(\pi) = p = \operatorname{nrd}(\alpha)\operatorname{nrd}(\beta)$ so either $\operatorname{nrd}(\alpha) = 1$ or $\operatorname{nrd}(\beta) = 1$, thus $\alpha \in \mathcal{O}^\times$ or $\beta \in \mathcal{O}^\times$. Conversely, suppose π is irreducible and let $p \mid \operatorname{nrd}(\pi)$. Let $I = \pi\mathcal{O} + p\mathcal{O} = \alpha\mathcal{O}$. Then $\operatorname{nrd}(\alpha) \mid \operatorname{nrd}(p) = p^2$. We cannot have $\operatorname{nrd}(\alpha) = 1$, as every element of I has reduced norm divisible by p. We similarly cannot have $\operatorname{nrd}(\alpha) = p^2$, since this would imply $\pi \in p\mathcal{O}$; but by Lemma 11.4.1, p is reducible, a contradiction. We conclude that $\operatorname{nrd}(\alpha) = p$. From $\pi \in I = \alpha\mathcal{O}$ we obtain $\pi = \alpha\beta$ with $\beta \in \mathcal{O}$; by irreducibility, $\beta \in \mathcal{O}^\times$ and $\operatorname{nrd}(\pi) = \operatorname{nrd}(\alpha) = p$. □

Definition 11.4.7. An element $\alpha \in \mathcal{O}$ is **primitive** if $\alpha \notin n\mathcal{O}$ for all $n \in \mathbb{Z}_{\geq 2}$.

Theorem 11.4.8. (Conway–Smith). *Let $\alpha \in \mathcal{O}$ be primitive and let $a = \operatorname{nrd}(\alpha)$. Factor $a = p_1 p_2 \cdots p_r$ into a product of primes. Then there exists $\pi_1, \pi_2, \dots, \pi_r \in \mathcal{O}$ such that*

$$\alpha = \pi_1\pi_2\cdots\pi_r, \quad \textit{and } \operatorname{nrd}(\pi_i) = p_i \textit{ for all } i. \tag{11.4.9}$$

Moreover, every other such factorization is of the form

$$\alpha = (\pi_1\gamma_1)(\gamma_1^{-1}\pi_2\gamma_2)\cdots(\gamma_{r-1}^{-1}\pi_r) \tag{11.4.10}$$

where $\gamma_1, \dots, \gamma_r \in O^\times$.

Proof. Let $I = \alpha O + p_1 O$; as in the proof of Lemma 11.4.6, we find $I = \pi_1 O$ with $\mathrm{nrd}(\pi_1) = p_1$, arguing that $\mathrm{nrd}(\pi_1) \neq p_1^2$ since $\alpha \in p_1 O$ is in contradiction to α being primitive. Then π_1 is unique up to right multiplication by a unit and $\alpha = \pi_1\alpha_2$. The result then follows by induction. □

The factorization (11.4.10) is said to be obtained from $\alpha = \pi_1 \cdots \pi_r$ by **unit migration**.

Remark 11.4.11. To look at all possible prime factorizations of α as in (11.4.9), it is necessary to consider the possible factorizations $a = p_1 \cdots p_r$. Conway–Smith call this process *metacommutation* [CSm2003, Chapter 5]; metacommutation is analyzed by Cohn–Kumar [CK2015], Forsyth–Gurev–Shrima FGS, and in a very general context by Chari [Cha2020].

11.5 Finite quaternionic unit groups

We conclude this section by a discussion of quaternion unit groups extending the discussion 11.2: we classify finite subgroups of $\mathbb{H}^\times$ and realize the possible subgroups as coming from quaternionic unit groups.

11.5.1. To begin with the classification, suppose that $\Gamma \subseteq \mathbb{H}^\times$ is a finite subgroup. Then $\mathrm{nrd}(\Gamma)$ is a finite subgroup of $\mathbb{R}^\times_{>0}$, hence identically 1, so $\Gamma \subseteq \mathbb{H}^1$.

Similarly, if $\Gamma \subseteq \mathbb{H}^\times/\mathbb{R}^\times \simeq \mathbb{H}^1/\{\pm 1\}$ is a finite subgroup, then it lifts via the projection $\mathbb{H}^1 \to \mathbb{H}^1/\{\pm 1\}$ to a finite subgroup of $\mathbb{H}^1$.

So let $\Gamma \subseteq \mathbb{H}^1$ be a finite subgroup. Then

$$\Gamma/\{\pm 1\} \hookrightarrow \mathbb{H}^1/\{\pm 1\} \simeq \mathrm{SO}(3)$$

the latter isomorphism by Hamilton's original (!) motivation for quaternion algebras (Corollary 2.4.21). Therefore $\Gamma/\{\pm 1\} \subseteq \mathrm{SO}(3)$ is a finite rotation group, and these groups have been known since antiquity.

Proposition 11.5.2. *A finite subgroup of* SO(3) *is one of the following:*

(i) *a cyclic group;*
(ii) *a dihedral group;*
(iii) *the tetrahedral group* A_4 *of order* 12*;*

(iv) *the octahedral group* S_4 *of order* 24*; or*
(v) *the icosahedral group* A_5 *of order* 60*.*

Cases (iii)–(v) are the symmetry groups of the corresponding Platonic solids and are called **exceptional** rotation groups.

Proof. Let $G \leq \mathrm{SO}(3)$ be a finite subgroup with $\#G = n > 1$; then G must consist of rotations about a common fixed point (its center of gravity), which we may take to be the origin. The group G then acts on the unit sphere, and every nonidentity element of G acts by rotation about an axis, fixing the poles of its axis on the sphere. Let V be the subset of these poles in the unit sphere; the set V will soon be the vertices of our (possibly degenerate) polyhedron. Let

$$X = \{(g, v) : g \in G \smallsetminus \{1\} \text{ and } v \text{ is a pole of } g\}.$$

Since each $g \in G \smallsetminus \{1\}$ has exactly two poles, we have $\#X = 2(n-1)$. On the other hand, we can also count organizing by orbits. Choose a representative set $v_1, \dots, v_r$ of poles, one from each orbit of G on V, and let

$$n_i = \#\operatorname{Stab}_G(v_i) = \#\{g \in G : gv_i = v_i\}$$

be the order of the stabilizer: this group is a cyclic subgroup about a common axis. Then

$$2n - 2 = \#X = \sum_{i=1}^{r} \#(Gv_i)(n_i - 1) = \sum_{i=1}^{r} \frac{n}{n_i}(n_i - 1) = n \sum_{i=1}^{r} \left(1 - \frac{1}{n_i}\right),$$

by the orbit–stabilizer theorem. Dividing both sides by n gives

$$2 - \frac{2}{n} = \sum_{i=1}^{r} \left(1 - \frac{1}{n_i}\right). \tag{11.5.3}$$

Since $n > 1$, we have $1 \leq 2 - 2/n < 2$; and since each $n_i \geq 2$, we have $1/2 \leq 1 - 1/n_i < 1$. Putting these together, we must have $r = 2, 3$.

If $r = 2$, then (11.5.3) becomes $2 = n/n_1 + n/n_2$, with $n/n_i = \#(Gv_i) \geq 1$, so $n_1 = n_2 = n$, there is only one axis of rotation, and G is cyclic.

If $r = 3$, then the only possibilities for (n_1, n_2, n_3) with $n_1 \leq n_2 \leq n_3$ are $(2, 2, c), (2, 3, 3), (2, 3, 4), (2, 3, 5)$; the corresponding groups have sizes $2c, 12, 24, 60$, respectively, and can be identified with D_{2c}, A_4, S_4, A_5 by a careful but classical analysis of orbits. See Armstrong [Arm88, Chapter 19], Grove–Benson [GB2008, §2.4], or Conway–Smith [CSm2003, §3.3]. □

In 11.2.4, we gave a quaternionic visualization of the binary tetrahedral group (lifting the tetrahedral group to $\mathbb{H}^1$); we repeat this with the two other exceptional rotation groups, taking again $k := ij$.

11.5.4. The octahedral group S_4 pulls back to the **binary octahedral group** $2O \subseteq \mathbb{H}^1$ of order $24 \cdot 2 = 48$, whose elements act by rigid motions of the octahedron (or dually, the cube). We make identifications following 11.2.4, shown in Figure 11.5.5.

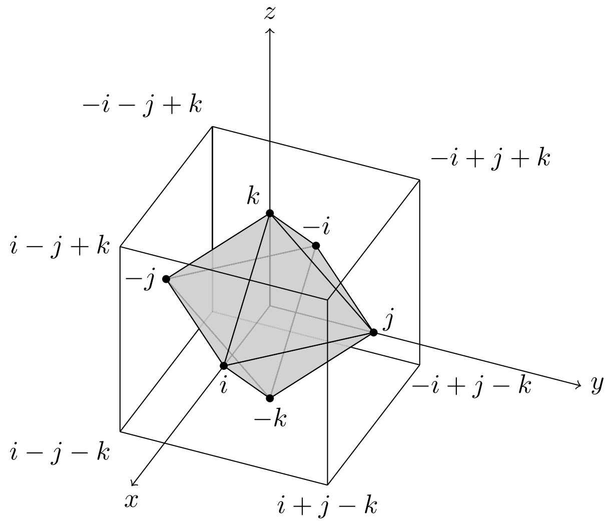


Figure 11.5.5: Symmetries of an octahedron and a cube, viewed quaternionically

The binary tetrahedral group $2T \trianglelefteq 2O$ of order 24 acts as a subgroup of rigid motions; the group $2O$ is generated by an element which maps to a rotation of order 4 around the 6 faces, i.e., one of the 12 elements

$$\frac{\pm 1 \pm i}{\sqrt{2}}, \frac{\pm 1 \pm j}{\sqrt{2}}, \frac{\pm 1 \pm k}{\sqrt{2}}.$$

The group $2O$ has a Coxeter presentation

$$2O \simeq \langle r, s, t \mid r^2 = s^3 = t^4 = rst = -1 \rangle$$

(with -1 central and $(-1)^2 = 1$). One also writes $2O \simeq \widetilde{S}_4$.

Let $F = \mathbb{Q}(\sqrt{2})$ and $R = \mathbb{Z}[\sqrt{2}]$. If we consider the Hamiltonians restricted to F as $B = \left(\dfrac{-1,-1}{F}\right)$, then the group $2O \subseteq \mathbb{H}^1$ generates an R-order: letting i, j be the standard generators and still $k := ij$, and letting $\alpha = (1+i)/\sqrt{2}$ and $\beta = (1+j)/\sqrt{2}$, then

$$O_{2O} = R + R\alpha + R\beta + R\alpha\beta; \tag{11.5.6}$$

this order contains the scalar extension of the Hurwitz order to R and is in fact a maximal R-order. (The extension of scalars is necessary: S_4 contains an element of order 4 which lifts to an element of order 8 in $2O$; such an element has trace $\pm(\zeta_8 + \zeta_8^{-1}) = \pm\sqrt{2}$.)

11.5.7. Finally, we treat the **binary icosahedral group** $2I \subseteq \mathbb{H}^1$ of order $60 \cdot 2 = 120$, acting by rigid motions of the icosahedron (or dually, the dodecahedron). We choose the regular icosahedron to have vertices at

$$\pm i \pm j \pm k,\ \pm\tau i \pm \tau^{-1} j,\ \pm\tau j \pm \tau^{-1} k,\ \pm\tau k \pm \tau^{-1} i$$

where $\tau = (1+\sqrt{5})/2$ is the golden ratio. The elements of order 5 are given by conjugates and powers of the element $\zeta = (\tau + \tau^{-1} i + j)/2$, which acts by rotation about a face. The group $2I$ can be presented as

$$2I \simeq \langle r, s, t \mid r^2 = s^3 = t^5 = rst = -1 \rangle$$

and we have $2I \simeq \widetilde{A}_5 \simeq \mathrm{SL}_2(\mathbb{F}_5)$. Letting now $F = \mathbb{Q}(\sqrt{5})$ and $R = \mathbb{Z}[\tau]$, the R-algebra generated by $2I$ is the maximal order

$$O_{2I} = R + Ri + R\zeta + Ri\zeta. \tag{11.5.8}$$

For further references, see Conway–Sloane [CSl88, §8.2], who describe the binary icosahedral group in detail, calling it the **icosian group**.

We now consider the related possibilities over $\mathbb{Q}$. (We will return to a general classification in section 32.4.) To put ourselves in a situation like (11.2.6), let $B = \left(\dfrac{a,b}{\mathbb{Q}}\right)$ be a quaternion algebra over $\mathbb{Q}$ such that $B \otimes_{\mathbb{Q}} \mathbb{R} = \left(\dfrac{a,b}{\mathbb{R}}\right) \simeq \mathbb{H}$: in this case, we say that B is **definite**. By Exercise 2.4, B is definite if and only if $a, b < 0$. Let $O \subseteq B$ be an order in B; we would like to understand its unit group.

Lemma 11.5.9. *The group $O^\times = O^1$ is finite.*

Proof. We may take $B = (a, b \mid \mathbb{Q})$ with $a, b < 0$. Consider the reduced norm $\mathrm{nrd}\colon B \to \mathbb{Q}$, given by $\mathrm{nrd}(t+xi+yj+zij) = t^2+|a|x^2+|b|y^2+|ab|z^2$, so $\mathrm{nrd}(B^\times) \subseteq \mathbb{Q}^\times_{>0}$. At the same time, $\mathrm{nrd}(O^\times) \subseteq \mathbb{Z}^\times = \{\pm 1\}$, so we conclude $O^\times = O^1$. This group is finite because the restriction $\mathrm{nrd}|_O$ of the reduced norm to $O \simeq \mathbb{Z}^4$ defines a (still) positive definite quadratic form, so there are only finitely many elements of O of any fixed reduced norm. (For a geometric perspective, viewing the elements of O^1 as lattice points on an ellipsoid in $\mathbb{R}^4$), see Proposition 17.5.6.) ◻

In view of Lemma 11.5.9, the classification of finite rotation groups (Proposition 11.5.2) applies. We consider each case in turn.

11.5.10. Among the (nontrivial) cyclic groups, only subgroups of order $2, 4, 6$ are possible over $\mathbb{Q}$. Indeed, a generator satisfies a quadratic equation with integer coefficients and so belongs to the ring of integers of an imaginary quadratic field; and only two imaginary quadratic fields have units other than ± 1, namely, the Eisenstein order $\mathbb{Z}[(-1+\sqrt{-3})/2]$ of discriminant -3 and the Gaussian order $\mathbb{Z}[\sqrt{-1}]$ of discriminant -4 with groups of size $4, 6$, respectively. (The more precise question of whether or not there is a unit of specified order is a question of embedding numbers, the subject of Chapter 30.)

11.5.11. Next, suppose that $\mathcal{O}^\times/\{\pm 1\}$ is dihedral, and let $j \in \mathcal{O}^\times \smallsetminus \{\pm 1\}$ act by inversion (equivalently, conjugation) on a cyclic group (of order 2, 3, by 11.5.10), generated by an element i. Let $K = \mathbb{Q}(i)$. Since j acts by inversion, we have $j^2 \in \mathbb{Q}$, and since $j \in \mathcal{O}^\times$ we have $j^2 = -1$. It follows that $j\alpha = \overline{\alpha} j$ for all $\alpha \in K$. Thus $B \simeq \left(\dfrac{K, -1}{\mathbb{Q}}\right)$, and we have two possibilities:

(i) If i has order 4, then $B \simeq (-1, -1 \mid \mathbb{Q})$ and $\mathcal{O}$ contains the order generated by i, j. This is the case treated in section 11.2: $\mathcal{O}$ is the Lipschitz order, and $\mathcal{O}^\times \simeq Q_8$ is the quaternion group of order 8.

(ii) Otherwise, $i = \omega$ has order 6, and $B \simeq (-3, -1 \mid \mathbb{Q})$. By Exercise 11.11(a), we have $(-3, -1 \mid \mathbb{Q}) \not\simeq (-1, -1 \mid \mathbb{Q})$. By an argument similar to Lemma 11.1.2—and boy, there is more of this to come in Chapter 32—we see that

$$\mathcal{O} = \mathbb{Z} + \mathbb{Z}\omega + \mathbb{Z}j + \mathbb{Z}\omega j \tag{11.5.12}$$

is maximal. The group $\mathcal{O}^\times/\{\pm 1\} \simeq D_6$ is a dihedral group of order 6, and the group $\mathcal{O}^\times$ is generated by ω, j with relations $\omega^3 = j^2 = -1$ and $j\omega = \omega^{-1} j$; in other words, $\mathcal{O}^\times \simeq C_3 \rtimes C_4$ is the semidirect product of the cyclic group C_3 of order 3 by the action of the cyclic group C_4 with a generator acting by inversion on C_3. Because $i^2 = -1$ is central, we also have an exact sequence

$$1 \to C_6 \to \mathcal{O}^\times \to C_2 \to 1$$

where $C_6 \simeq \langle \omega \rangle$ and $C_2 \simeq \langle j \rangle/\{\pm 1\}$. This group is also called the **binary dihedral** or **dicyclic** group of order 12, denoted $2D_6$.

11.5.13. To conclude, suppose that $\mathcal{O}^\times/\{\pm 1\}$ is exceptional. Each of these groups contain a dihedral group, so the argument from 11.5.11 applies: the only new group we see is the (binary) tetrahedral group obtained from the Hurwitz units (section 11.2). Here is another proof: the group S_4 contains an element of order 4 and A_5 an element of order 5, and these lift to elements of order 8, 10 in $\mathcal{O}^\times$, impossible.

We have proven the following theorem.

Theorem 11.5.14. *Let $B = (a, b \mid \mathbb{Q})$ be a quaternion algebra over $\mathbb{Q}$ with $a, b < 0$, and let $\mathcal{O} \subseteq B$ be an order. Then $\mathcal{O}^\times$ is either cyclic of order* 2, 4, 6*, quaternion of order* 8*, binary dihedral of order* 12*, or binary tetrahedral of order* 24*.*

Moreover, $\mathcal{O}^\times$ is quaternion, binary dihedral, or binary tetrahedral if and only if $\mathcal{O}$ is isomorphic to the Lipschitz order, the order (11.5.12)*, or the Hurwitz order, respectively.*

Proof. Combine 11.5.10, 11.5.11, and 11.5.13. □

Exercises

▷ 1. Check directly that the Hurwitz order

$$\mathcal{O} = \mathbb{Z} + \mathbb{Z}i + \mathbb{Z}j + \mathbb{Z}\left(\frac{1+i+j+k}{2}\right)$$

is indeed an order in $B = \left(\frac{-1,-1}{\mathbb{Q}}\right)$.

2. Let $B = \left(\frac{-1,-1}{\mathbb{Q}}\right)$, and let $\mathcal{O} \subseteq B$ be the Hurwitz order. For the normalizer

$$N_{B^\times}(\mathcal{O}) := \{\alpha \in B^\times : \alpha^{-1}\mathcal{O}\alpha = \mathcal{O}\}$$

show the equality $N_{B^\times}(\mathbb{Z}\langle i, j\rangle) = N_{B^\times}(\mathcal{O})$. *[Hint: consider units and their traces.]*

3. (a) Show that the Lipschitz order $\mathbb{Z}\langle i, j\rangle$ is the *unique* suborder of the Hurwitz order $\mathcal{O}$ with index 2 (as abelian groups).

   (b) Show that
$$\mathbb{Z}\langle i, j\rangle = \{\alpha \in \mathcal{O} : \operatorname{trd}(\alpha) \text{ is even}\}.$$

Check that the map

$$\begin{aligned} \mathcal{O}/3\mathcal{O} &\to \mathrm{M}_2(\mathbb{F}_3) \\ i, j &\mapsto \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}, \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix} \end{aligned}$$

from Lemma 11.2.1 is an $\mathbb{F}_3$-algebra isomorphism.

4. Generalizing the previous exercise, show that for an odd prime p that $\mathcal{O}/p\mathcal{O} \simeq \mathrm{M}_2(\mathbb{F}_p)$.

5. Draw the subgroup lattice for $\mathrm{SL}_2(\mathbb{F}_3)$, indicating normal subgroups (and their quotients).

▷ 6. Show explicitly that

$$2T \simeq \langle r, s, t \mid r^2 = s^3 = t^3 = rst = -1\rangle$$

(cf. (11.2.8)).

▷ 7. Let
$$\Lambda = \mathbb{Z}^4 + \mathbb{Z}(\tfrac{1}{2}, \tfrac{1}{2}, \tfrac{1}{2}, \tfrac{1}{2}) \subset \mathbb{R}^4$$

be the image of the Hurwitz order $\mathcal{O}$ under the natural embedding $\mathcal{O} \hookrightarrow \mathbb{H} \simeq \mathbb{R}^4$. Show that for every $x \in \mathbb{R}^4$, there exists $\lambda \in \Lambda$ such that $\|\lambda\|^2 \leq 1/2$. *[Hint: without loss of generality we may take $0 \leq x_i \leq 1/2$ for all i; then show we may take $x_1 + x_2 + x_3 + x_4 \leq 1$; conclude that the maximum value of $\|x\|^2$ with these conditions occurs at the point $(\frac{1}{2}, \frac{1}{2}, 0, 0)$.]*

8. Let B be a definite quaternion algebra over $\mathbb{Q}$ and let $\mathcal{O} \subseteq B$ be an order. Show that $\mathcal{O}$ is left Euclidean if and only if $\mathcal{O}$ is right Euclidean (with respect to a norm N).

9. Let $\mathcal{O} \subset B := (-1, -1 \mid \mathbb{Q})$ be the Hurwitz order.

   (a) Consider the natural ring homomorphism $\mathcal{O} \to \mathcal{O}/2\mathcal{O} = \mathcal{O} \otimes_{\mathbb{Z}} \mathbb{F}_2$ giving the reduction of the algebra $\mathcal{O}$ modulo 2. Show that $\mathcal{O}/2\mathcal{O}$ is an $\mathbb{F}_2$-algebra, that $\#(\mathcal{O}/2\mathcal{O}) = 16$, and that $(\mathcal{O}/2\mathcal{O})^\times \simeq A_4$ is isomorphic to the alternating group on 4 elements. Conclude that $\mathcal{O}/2\mathcal{O} \not\simeq \mathrm{M}_2(\mathbb{F}_2)$ and hence that $\mathcal{O}/2\mathcal{O}$ is *not* a quaternion algebra over $\mathbb{F}_2$.

(b) Show that the group of ring automorphisms of $O/2O$ is
$$\mathrm{Aut}(O/2O) \simeq S_4.$$

(c) More generally, if F is a field of characteristic 2 show that there is an exact sequence
$$1 \to F^2 \to \mathrm{Aut}(O \otimes_{\mathbb{Z}} F) \to \mathrm{SL}_2(F) \to 1$$
where F^2 is an additive group. *[Hint: let $J = \mathrm{rad}(O \otimes_{\mathbb{Z}} F)$ be the Jacobson radical of the algebra, and show that the sequence is induced by F-linear automorphisms of J and the automorphisms $\omega \mapsto \omega + \epsilon$* with *$\epsilon \in J$.]*

[This kind of construction, considered instead over the octonions, arises when constructing the exceptional group G_2 in characteristic 2 [Wils2009, §4.4.1].]

10. Although the Lipschitz order just misses being Euclidean with respect to the norm (see Example 11.3.8), bootstrapping from the Hurwitz order we still obtain a result on principality by restricting the set of ideals, as follows.

Let $I \subseteq \mathbb{Z}\langle i, j\rangle$ be a right ideal.

(a) Show that $IO = \beta O$ for some $\beta \in IO \cap \mathbb{Z}\langle i, j\rangle$.
(b) Prove that $I_{(2)} = \mathbb{Z}_{(2)}\langle i, j\rangle$ if and only if I is generated by elements of odd reduced norm.
(c) If $I_{(2)} = \mathbb{Z}_{(2)}\langle i, j\rangle$, show that $IO \cap \mathbb{Z}\langle i, j\rangle = I$ and conclude that I is right principal. *[Hint: Argue locally.]*

11. Let $B := (-1, -3 \mid \mathbb{Q})$, and let
$$O := \mathbb{Z}\langle i, (1 + j)/2\rangle = \mathbb{Z} + \mathbb{Z}i + \mathbb{Z}\frac{1 + j}{2} + \mathbb{Z}i\frac{1 + j}{2}.$$

(a) Show that $B \not\simeq (-1, -1 \mid \mathbb{Q})$.
(b) Show that O is a maximal order in B.
(c) Show that O is Euclidean with respect to the reduced norm
$$\mathrm{nrd}(t + xi + y(1 + j)/2 + zi(1 + j)/2) = t^2 + ty + x^2 + xz + y^2 + z^2.$$
(d) Show that every maximal order in B is conjugate to O.

12. Let $G \leq \mathrm{O}(2)$ be a finite subgroup such that $\mathrm{tr}(g), \det(g) \in \mathbb{Q}$ for all $g \in G$. Show that G is conjugate in $\mathrm{O}(2)$ to one of the following: (i) a cyclic group of order $1, 2, 3, 4, 6$ that is a subgroup of $\mathrm{SO}(2)$, or (ii) a dihedral group of order $2, 4, 6, 8, 12$, not contained in $\mathrm{SO}(2)$.

13. Let p be an odd prime.

(a) The group $\mathrm{GL}_2(\mathbb{Z}_{(p)})$ acts by right multiplication on the set of matrices $\pi \in \mathrm{M}_2(\mathbb{Z}_{(p)})$ with $p \parallel \det(\pi)$ (i.e., p exactly divides the numerator of $\det(\pi)$, written in lowest terms). Show that there are precisely $p + 1$ orbits, represented by

$$\pi = \begin{pmatrix} p & 0 \\ 0 & 1 \end{pmatrix}$$

and

$$\pi = \begin{pmatrix} 1 & 0 \\ a & p \end{pmatrix}, \quad a = 0, 1, \dots, p-1.$$

[Hint: use column operations.]

(b) Repeat (a) but with $\mathrm{SL}_2(\mathbb{Z}_{(p)})$ acting on the set of matrices $\pi \in \mathrm{M}_2(\mathbb{Z}_{(p)})$ with $\det(\pi) = p$, with the same conclusion.

(c) Show that the number of (left or) right ideals of $\mathcal{O}$ of reduced norm p is equal to $p + 1$.

(d) Accounting for units, conclude that the number of ways of writing an odd prime p as the sum of four squares is equal to $8(p + 1)$.

14. In the following exercise, we consider a computational problem, suitable for those with some background in number theory algorithms (see e.g. Cohen [Coh93]).

(a) Show that one can find $x, y, z \in \mathbb{Z}$ such that $x^2 + y^2 + z^2 = pm$ with $p \nmid m$ in probabilistic polynomial time in $\log p$.

(b) Describe the right Euclidean algorithm as applied to $\alpha = xi + yj + zk$ and p to obtain $\pi \in \mathcal{O}$ with $\operatorname{nrd}(\pi) = p$. Adjust as in Lemma 11.2.9 to find a solution to $x^2 + y^2 + z^2 = p$ with $x, y, z \in \mathbb{Z}$. Estimate the running time of this algorithm.

This chapter is licensed under the terms of the Creative Commons Attribution-NonCommercial 4.0 International License (http://creativecommons.org/licenses/by-nc/4.0/), which permits any noncommercial use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license and indicate if changes were made.

The images or other third party material in this chapter are included in the chapter's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the chapter's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.

Chapter 12

Ternary quadratic forms over local fields

In this chapter, we classify quaternion algebras over local fields using quadratic forms; this generalizes the classification of quaternion algebras over $\mathbb{R}$.

12.1 ▹ The p-adic numbers and local quaternion algebras

Before beginning, we briefly remind the reader about the structure of the p-adic numbers. The p-adics were developed by Hensel, who wanted a uniform way to say that a Diophantine equation has a (consistent) solution modulo p^n for all n. In the early 1920s, Hasse used them in the study of quadratic forms and algebras over number fields. At the time, what is now called the *local-global principle* then was called the p-adic transfer from the "small" to the "large". As references on p-adic numbers, see for example Gouvêa [Gou97], Katok [Kat2007], or Koblitz [Kob84].

Just as elements of $\mathbb{R}$ can be thought of infinite decimals, an element of $\mathbb{Q}_p$ can be thought of in its p-adic expansion

$$a = (\dots a_3 a_2 a_1 a_0 . a_{-1} a_{-2} \cdots a_{-k})_p = \sum_{n=-k}^{\infty} a_n p^n \tag{12.1.1}$$

where each $a_i \in \{0, \dots, p-1\}$ are the **digits** of a. We continue "to the left" because a decimal expansion is a series in the base $1/10 < 1$ and instead we have a base $p > 1$.

Put a bit more precisely, we define the **p-adic absolute value** on $\mathbb{Q}$ by defined by $|0|_p := 0$ and

$$|c|_p := p^{-v_p(c)} \quad \text{for } c \in \mathbb{Q}^\times, \tag{12.1.2}$$

where $v_p(c)$ is the power of p occurring in c in its unique factorization (taken to be negative if p divides the denominator of c written in lowest terms). Then the field $\mathbb{Q}_p$ is the completion of $\mathbb{Q}$ with respect to $|\,|_p$, that is to say, $\mathbb{Q}_p$ is the set of equivalence classes of Cauchy sequences of rational numbers, and it obtains a topology induced by the metric $d_p(x, y) = |x - y|_p$. We have $|a|_p = p^k$ for a as in (12.1.1) with $a_{-k} \neq 0$.

© The Author(s) 2021

J. Voight, *Quaternion Algebras*, Graduate Texts in Mathematics 288,
https://doi.org/10.1007/978-3-030-56694-4_12

Of course, all of the information in the p-adic absolute value is encoded in the p**-adic valuation** $v_p : \mathbb{Q} \to \mathbb{R} \cup \{\infty\}$.

Inside $\mathbb{Q}_p$ is the ring $\mathbb{Z}_p$ of p**-adic integers**, the completion of $\mathbb{Z}$ with respect to $|\,|_p$: the ring $\mathbb{Z}_p$ consists of those elements of $\mathbb{Q}_p$ with $a_n = 0$ for $n < 0$. (The ring $\mathbb{Z}_p$ might be thought of intuitively as $\mathbb{Z}/p^\infty\mathbb{Z}$, if this made sense.)

Equipped with their topologies, the ring $\mathbb{Z}_p$ is compact and the field $\mathbb{Q}_p$ is locally compact. These statements can be understood quite easily by viewing $\mathbb{Z}_p$ in a slightly different way, as a projective limit with respect to the natural projection maps $\mathbb{Z}/p^{n+1}\mathbb{Z} \to \mathbb{Z}/p^n\mathbb{Z}$:

$$\begin{aligned}\mathbb{Z}_p &= \varprojlim_n \mathbb{Z}/p^n\mathbb{Z} \\ &= \left\{ x = (x_n)_n \in \prod_{n=1}^{\infty} \mathbb{Z}/p^n\mathbb{Z} : x_{n+1} \equiv x_n \pmod{p^n} \text{ for all } n \geq 1 \right\}.\end{aligned} \tag{12.1.3}$$

In other words, each element of $\mathbb{Z}_p$ is a compatible sequence of elements in $\mathbb{Z}/p^n\mathbb{Z}$ for each n. The equality (12.1.3) is just a reformulation of the notion of Cauchy sequence for $\mathbb{Z}$, and so for the purposes of this introduction it can equally well be taken as a definition.

As for the topology in (12.1.3), each factor $\mathbb{Z}/p^n\mathbb{Z}$ is given the discrete topology, the product $\prod_{n=0}^{\infty} \mathbb{Z}/p^n\mathbb{Z}$ is given the product topology, and $\mathbb{Z}_p$ is given the subspace topology. Since each $\mathbb{Z}/p^n\mathbb{Z}$ is compact (it is a finite set!), by Tychonoff's theorem the product $\prod_{n=0}^{\infty} \mathbb{Z}/p^n\mathbb{Z}$ is compact; and $\mathbb{Z}_p$ is closed inside this product (a convergent limit of Cauchy sequences is a Cauchy sequence), so $\mathbb{Z}_p$ is compact and still Hausdorff. The topology on $\mathbb{Z}_p$ is a bit strange though, as $\mathbb{Z}_p$ is **totally disconnected**: every nonempty connected subset is a single point. In fact, $\mathbb{Z}_p$ is homeomorphic to the Cantor set, which is itself homeomorphic to the product of countably many copies of $\{0, 1\}$. (More generally, every nonempty totally disconnected compact metric space with no isolated points is homeomorphic to the Cantor set.)

The set $\mathbb{Z}_p$ is a compact neighborhood of 0, as it is the closed ball of radius 1 around 0:

$$\mathbb{Z}_p = \{x \in \mathbb{Q}_p : |x|_p \leq 1\} = \{x \in \mathbb{Q}_p : v_p(x) \geq 0\}. \tag{12.1.4}$$

In a similar way, the disc of radius 1 around $a \in \mathbb{Q}_p$ is a compact neighborhood of a homeomorphic to $\mathbb{Z}_p$, so $\mathbb{Q}_p$ is locally compact. Being able to make topological arguments like the one above is the whole point of looking at fields like $\mathbb{Q}_p$: our understanding of infinite algebraic objects is informed by topology.

With this review, and topological arguments now at our disposal, we consider quaternion algebras over $\mathbb{Q}_p$. The 'original' quaternion algebra, of course, was the division ring $\mathbb{H}$ of Hamiltonians over the real numbers (the 'original' field with a topology), and indeed $\mathbb{H}$ is the unique division quaternion algebra over $\mathbb{R}$ (Corollary 3.5.8). We find a similar result over $\mathbb{Q}_p$ (a special case of Theorem 12.3.2), as follows.

Theorem 12.1.5. *There is a unique division quaternion algebra B over $\mathbb{Q}_p$, up to isomorphism; if $p \neq 2$, then*

$$B \simeq \left(\frac{e, p}{\mathbb{Q}_p}\right)$$

where $e \in \mathbb{Z}$ is a quadratic nonresidue modulo p.

For example, if $p \equiv 3 \pmod 4$ then -1 is quadratic nonresidue and $(-1, p \mid \mathbb{Q}_p)$ is the unique division quaternion algebra over $\mathbb{Q}_p$.

Because we have exactly two such possibilities, we define the **Hilbert symbol**: for $a, b \in \mathbb{Q}_p^\times$, we have $(a, b)_{\mathbb{Q}_p} = 1, -1$ according as the quaternion algebra $(a, b \mid \mathbb{Q}_p) \simeq \mathrm{M}_2(\mathbb{Q}_p)$ is split or not. According to Theorem 12.1.5, the Hilbert symbol over $\mathbb{Q}_p$ uniquely identifies the two possible isomorphism classes of quaternion algebras over $\mathbb{Q}_p$—just like it does over $\mathbb{R}$.

Our approach to Theorem 12.1.5 uses quadratic forms: we use the classification of isomorphism classes of quaternion algebras given in terms of similarity classes of ternary quadratic forms (Theorem 5.1.1). The following proposition then implies Theorem 12.1.5.

Proposition 12.1.6. *There is a unique ternary anisotropic quadratic form Q over $\mathbb{Q}_p$, up to similarity; if $p \neq 2$, then $Q \sim \langle 1, -e, -p\rangle$ where e is a quadratic nonresidue modulo p.*

Happily, this proposition can be proved using some rather direct manipulations with quadratic forms and gives a very "hands on" feel; it is also suggests the arguments we use for a more general result. The main input we need is a quadratic Hensel's lemma, or more precisely, the following consequence.

Lemma 12.1.7. *For $p \neq 2$, the classes in $\mathbb{Q}_p^\times/\mathbb{Q}_p^{\times 2}$ are represented by $1, e, p, ep$ where e is a quadratic nonresidue modulo p.*

Proof. Let $a \in \mathbb{Q}_p^\times$ and let $m := v_p(a)$. Then $a = bp^m$ with $b := a/p^m \in \mathbb{Z}_p^\times$, and by squaring $a \in \mathbb{Q}_p^{\times 2}$ if and only if $b \in \mathbb{Z}_p^{\times 2}$ and m is even. We claim that $b \in \mathbb{Z}_p^{\times 2}$ if and only if its reduction b modulo p is a square in $(\mathbb{Z}/p\mathbb{Z})^\times$. With the forward implication immediate, suppose $b \equiv c^2 \pmod p$ with $c \in \mathbb{Z}_p^\times$, then $b/c^2 \in 1 + p\mathbb{Z}_p$. But squaring is a bijection on $1 + p\mathbb{Z}_p$, by expanding the square root as a convergent series (see Exercise 12.1) and using that $p \neq 2$. Thus $b/c^2 \in \mathbb{Z}_p^{\times 2}$, and the result follows. □

We now proceed with the proof when $p \neq 2$.

Proof of Proposition 12.1.6, $p \neq 2$. We start by showing that $Q(x, y, z) = x^2 - ey^2 - pz^2$ is anisotropic. Suppose $Q(x, y, z) = 0$ with not all $x, y, z \in \mathbb{Q}_p$ zero. Rescaling by p, we may assume that $x, y, z \in \mathbb{Z}_p$ and not all $x, y, z \in p\mathbb{Z}_p$. We then reduce modulo p to find that $x^2 \equiv ey^2 \pmod p$. If $p \nmid y$, then $(x/y)^2 \equiv e \pmod p$; but e is a quadratic nonresidue modulo p, a contradiction. So $p \mid y$; thus $p \mid (ey^2 + pz^2) = x^2$, so $p \mid x$; thus $p^2 \mid (x^2 - ey^2) = pz^2$, so $p \mid z$, a contradiction.

To show uniqueness, let Q be a ternary anisotropic form over $\mathbb{Q}_p$. Since $p \neq 2$, we may diagonalize. In such a diagonal form taken up to similarity, we may also rescale each coordinate up to squares as well as rescale the entire quadratic form. Putting this together with Lemma 12.1.7, without loss of generality we may suppose $Q(x, y, z) = \langle 1, -b, -c\rangle = x^2 - by^2 - cz^2$ with $b, c \in \{1, e, p, ep\}$, the signs chosen for convenience. If $b = 1$ or $c = 1$, the form is isotropic by inspection. So we are left to consider the cases $(b, c) = (e, e), (e, p), (e, ep), (p, p), (p, ep), (ep, ep)$.

- When $(b, c) = (e, e)$, we have after rescaling $x^2 + y^2 - ez^2$. We claim this form is always isotropic. Indeed, the form reduces to a nondegenerate ternary quadratic form over $\mathbb{F}_p$. Such a form is always isotropic by a delightful counting argument (Exercise 5.5(b), or a second chance in Exercise 12.6!). Lifting, there exist $x, y, z \in \mathbb{Z}_p$, not all zero modulo p, such that $x^2 \equiv -y^2 + ez^2 \pmod{p}$. Since e is a nonsquare, we have $p \nmid x$ (arguing similarly as in the first paragraph). Let $d := -y^2 + ez^2 \in \mathbb{Q}_p^\times$. By the possibilities in Lemma 12.1.7, we must have $d \in \mathbb{Q}_p^{\times 2}$; solving $x^2 = d$ for $x \in \mathbb{Q}_p$ then shows that Q is isotropic.
- The case (e, p) is our desired form.
- In the third case (eep!), we substitute $x \leftarrow ex$ and divide by e to obtain the form $-y^2 + ex^2 - pz^2$. We claim that there is an isometry $\langle -1, e\rangle \simeq \langle 1, -e\rangle$: indeed, in the first bullet we showed that the quadratic form $\langle -1, e, -1\rangle$ is isotropic, so $-x^2 + ey^2$ represents 1; using this representation as the first basis vector, extending to a basis, and diagonalizing, we conclude that $\langle -1, e\rangle \simeq \langle 1, b\rangle$. By discriminants, we have $-e = b$ up to squares. This brings us back to the first case.
- In cases (p, p) or (ep, ep), replacing $x \leftarrow px$ and dividing gives the quadratic forms $x^2 + y^2 - pz^2$ and $x^2 + y^2 - epz^2$. If $-1 \in \mathbb{Z}_p^{\times 2}$, then the form is isotropic; otherwise, we may take $e = -1$ and we are back to cases $(e, p), (e, ep)$.
- In the final case (keeping pep!), we substitute $x \leftarrow px$ and divide by $-p$ to get $y^2 + ez^2 - px^2$. If $-1 \in \mathbb{Z}_p^{\times 2}$, then by substitution we change the middle sign to return to the first case. Otherwise, we may take $e = -1$, and the form is isotropic, a contradiction.

This consideration of cases completes the proof. □

Although direct, the proof we just gave has the defect that quadratic forms behave differently in characteristic 2, and so one may ask for a proof that works uniformly in all characteristics: we give such a proof in the next chapter by extending valuations.

One of the nice applications of this classification is that it gives a *necessary* condition for two quaternion algebras to be isomorphic. Let $B = (a, b \mid \mathbb{Q})$ be a quaternion algebra over $\mathbb{Q}$ and consider its scalar extension $B_p = B \otimes_{\mathbb{Q}} \mathbb{Q}_p \simeq (a, b \mid \mathbb{Q}_p)$. If B' is another quaternion algebra over $\mathbb{Q}$ and $B \simeq B'$, then this implies $B_p \simeq B'_p$ for all primes p, and of course the same is true over $\mathbb{R}$. Perhaps surprisingly, it turns out that the collection of all of these tests is also sufficient: if B, B' become isomorphic over $\mathbb{R}$ and over $\mathbb{Q}_p$ for all primes p, then in fact $B \simeq B'$ are isomorphic over $\mathbb{Q}$! This profound and powerful principle—detecting global isomorphism from local isomorphisms, a *local-global principle*—will be examined in chapter 14.

12.2 Local fields

In this section, we set up notation and basic results from the theory of local fields. The theory of local fields is described in many places, including Neukirch [Neu99, Chapters II, V], the classic texts by Cassels [Cas86] and Serre [Ser79]. Weil [Weil74] approaches number theory from the ground up in the language of local fields, building up the theory of local division rings.

Our motivation for local fields is as follows: we want a topology compatible with the field operations in which the field is Hausdorff and locally compact (every element has a compact neighborhood), analogous to what holds over the real and complex numbers. And to avoid trivialities, we will insist that this topology is not the discrete topology (where every subset of F is open). To carry this out, we begin with some basic definitions.

Definition 12.2.1. A **topological group** is a group equipped with a topology such that the group operation and inversion are continuous. A **homomorphism** of topological groups is a group homomorphism that is continuous.

A **topological ring** is a ring A equipped with a topology such that the ring operations (addition, negation, and multiplication) are continuous. A **homomorphism** of topological rings is a ring homomorphism that is continuous. A **topological field** is a field that is also a topological ring in such a way that division by a nonzero element is continuous.

One natural way to equip a ring with a topology is by way of an absolute value. To get started, we consider such notions first for fields. Throughout this section, let F be a field.

Definition 12.2.2. An **absolute value** on F is a map

$$||: F \to \mathbb{R}_{\geq 0}$$

such that:

(i) $|x| = 0$ if and only if $x = 0$;
(ii) $|xy| = |x||y|$ for all $x, y \in F$; and
(iii) $|x + y| \leq |x| + |y|$ for all $x, y \in F$ (**triangle inequality**).

An absolute value $||$ on F gives F the structure of a topological field by the metric $d(x, y) = |x - y|$. Two absolute values $||_1, ||_2$ on F are **(strictly) equivalent** if there exists $c > 0$ such that $|x|_1 = |x|_2^c$ for all $x \in F$; equivalent absolute values induce the same topology on F.

Remark 12.2.3. If $||$ is an absolute value on F, then it need not be the case that $x \mapsto |x|^c$ for $c > 0$ is again absolute value, because it need not satisfy the triangle inequality. In particular, we will find it convenient to consider the *square* of the usual absolute value on $F = \mathbb{C}$, which suffers from this deficiency. There are various ways around this problem; perhaps the simplest is just to ignore it.

Definition 12.2.4. An absolute value is **nonarchimedean** if the **ultrametric inequality**

$$|x + y| \leq \sup\{|x|, |y|\}$$

is satisfied for all $x, y \in F$, and **archimedean** otherwise.

Example 12.2.5. The fields $\mathbb{R}$ and $\mathbb{C}$ are topological fields with respect to the usual archimedean absolute value.

Remark 12.2.6. A field with absolute value is archimedean if and only if it satisfies the **archimedean property**: for all $x \in F^\times$, there exists $n \in \mathbb{Z}$ such that $|nx| > 1$. In particular, a field F equipped with an archimedean absolute value has $\operatorname{char} F = 0$.

Example 12.2.7. Every field has the **trivial** (nonarchimedean) absolute value, defined by $|0| = 0$ and $|x| = 1$ for all $x \in F^\times$; the trivial absolute value induces the discrete topology on F.

A nonarchimedean absolute value on a field F arises naturally by way of a valuation, as follows.

Definition 12.2.8. A **valuation** of a field F is a map $v\colon F \to \mathbb{R} \cup \{\infty\}$ such that:

(i) $v(x) = \infty$ if and only if $x = 0$;
(ii) $v(xy) = v(x) + v(y)$ for all $x, y \in F$; and
(iii) $v(x + y) \geq \min(v(x), v(y))$ for all $x, y \in F$.

A valuation is **discrete** if the **value group** $v(F^\times)$ is discrete in $\mathbb{R}$ (has no accumulation points).

Here, we set the convention that $x + \infty = \infty + x = \infty$ for all $x \in \mathbb{R} \cup \{\infty\}$. By (ii), the value group $v(F^\times)$ is a subgroup of the additive group $\mathbb{R}$, and so whereas an absolute value is multiplicative, a valuation is additive.

Example 12.2.9. Each $x \in \mathbb{Q}^\times$ can be written $x = p^r a/b$ with $a, b \in \mathbb{Z}$ relatively prime and $p \nmid ab$; the map $v_p(x) = r$ defines the p-adic valuation on $\mathbb{Q}$.

Example 12.2.10. Let k be a field and $F = k(t)$ the field of rational functions over k. For $f(t) = g(t)/h(t) \in k(t) \smallsetminus \{0\}$ with $g(t), h(t) \in k[t]$, define $v(f(t)) := \deg h(t) - \deg g(t)$. Then v is a discrete valuation on F.

Given the parallels between them, it should come as no surprise that a valuation gives rise to an absolute value on F by defining

$$|x| = c^{-v(x)} \tag{12.2.11}$$

for a fixed $c > 1$; the induced topology on F is independent of the choice of c. By condition (iii), the absolute value associated to a valuation is nonarchimedean.

Example 12.2.12. The **trivial** valuation is the valuation v satisfying $v(0) = \infty$ and $v(x) = 0$ for all $x \in F^\times$. The trivial valuation gives the trivial absolute value on F.

Two valuations v, w are **equivalent** if there exists $a \in \mathbb{R}_{>0}$ such that $v(x) = aw(x)$ for all $x \in F$; equivalent valuations give the same topology on a field. A nontrivial discrete valuation is equivalent after rescaling (by the minimal positive element in the value group) to one with value group $\mathbb{Z}$, since a nontrivial discrete subgroup of $\mathbb{R}$ is cyclic; we call such a discrete valuation **normalized**.

12.2.13. Given a field F with a nontrivial discrete valuation v, the **valuation ring** is $R := \{x \in F : v(x) \geq 0\}$. We have $R^\times = \{x \in F : v(x) = 0\}$ since

$$v(x) + v(x^{-1}) = v(xx^{-1}) = v(1) = 0$$

for all $x \in F^\times$. The valuation ring is a local domain with unique maximal ideal

$$\mathfrak{p} := \{x \in F : v(x) > 0\} = R \smallsetminus R^\times.$$

An element $\pi \in \mathfrak{p}$ with smallest valuation is called a **uniformizer**, and comparing valuations we see that $\pi R = (\pi) = \mathfrak{p}$. Since $\mathfrak{p} \subsetneq R$ is maximal, the quotient $k := R/\mathfrak{p}$ is a field, called the **residue field** of R (or of F).

Recall that a topological space is **locally compact** if each point has a compact neighborhood (every point is contained in a compact set containing an open set).

Definition 12.2.14. A **local field** is a Hausdorff, locally compact topological field with a nondiscrete topology.

In a local field, we can hope to understand its structure by local considerations in a compact neighborhood, hence the name. Local fields have a very simple classification as follows.

Theorem 12.2.15. *A field F with absolute value is a local field if and only if F is one of the following:*

(i) *F is archimedean, and $F \simeq \mathbb{R}$ or $F \simeq \mathbb{C}$;*
(ii) *F is nonarchimedean with* char $F = 0$*, and F is a finite extension of $\mathbb{Q}_p$ for some prime p; or*
(iii) *F is nonarchimedean with* char $F = p$*, and F is a finite extension of the Laurent series field $\mathbb{F}_p((t))$ for some prime p; in this case, there is a (non-canonical) isomorphism $F \simeq \mathbb{F}_q((t))$ where q is a power of p.*

A field F with absolute value $|\ |$ is a nonarchimedean local field if and only if F is complete with respect to $|\ |$, and $|\ |$ is equivalent to the absolute value associated to a nontrivial discrete valuation $v \colon F \to \mathbb{R} \cup \{\infty\}$ with finite residue field.

Proof. See Neukirch [Neu99, Chapter II, §5], Cassels [Cas86, Chapter 4, §1], or Serre [Ser79, Chapter II, §1]. ∎

Although a local field is only locally compact, the valuation ring is itself compact, as follows.

Lemma 12.2.16. *Suppose F is nonarchimedean. Then F is totally disconnected and the valuation ring $R \subset F$ is a compact, totally disconnected topological ring.*

Proof. To see that F is totally disconnected (whence R too is totally disconnected), by translation it suffices to show that the only connected set containing 0 is $\{0\}$. Let

$x \in F^\times$ with $|x| = \delta > 0$. The image $|F^\times| \subseteq \mathbb{R}_{>0}$ is discrete, so there exists $0 < \epsilon < \delta$ such that $|y| < \delta$ implies $|y| \le \delta - \epsilon$ for all $y \in F$. Thus an open ball is a closed ball

$$D(0,\delta) = \{y \in F : |y| < \delta\} = \{y \in F : |y| \le \delta - \epsilon\} = D[0, \delta - \epsilon];$$

since $x \in F^\times$ and $\delta > 0$ were arbitrary, the only connected subset containing 0 is $\{0\}$.

Next, we show R is compact. There is a natural continuous ring homomorphism

$$\phi \colon R \to \prod_{n=1}^{\infty} R/\mathfrak{p}^n$$

where each factor $R/\mathfrak{p}^n$ is equipped with the discrete topology and the product is given the product topology. The map ϕ is injective, since $\bigcap_{n=1}^{\infty} \mathfrak{p}^n = \{0\}$ (every nonzero element has finite valuation). The image of ϕ is obviously closed. Therefore R is homeomorphic onto its closed image. But by Tychonoff's theorem, the product $\prod_{n=1}^{\infty} R/\mathfrak{p}^n$ of compact sets is compact, and a closed subset of a compact set is compact, thus R is compact. □

One key property of local fields we will use is Hensel's lemma.

Lemma 12.2.17. (Hensel's lemma, univariate). *Let F be a nonarchimedean local field with valuation v and valuation ring R, and let $f(x) \in R[x]$. Let $a \in R$ satisfy $m := v(f(a)) > 2v(f'(a))$. Then there exists $\widetilde{a} \in R$ such that $f(\widetilde{a}) = 0$ and $\widetilde{a} \equiv a \pmod{\mathfrak{p}^m}$.*

Proof. The result is straightforward to prove using Taylor expansion or the same formulas as in Newton's method. □

Perhaps less well-known is the multivariate version.

Lemma 12.2.18. (Hensel's lemma). *Let F be a nonarchimedean local field with valuation v and valuation ring R, and let $f(x_1, \dots, x_n) \in R[x_1, \dots, x_n]$ with $n \ge 1$.*

Let $a \in R^n$ have $m := v(f(a))$ and suppose that

$$m > 2 \min_i v\left(\frac{\partial f}{\partial x_i}(a)\right) \ge 0.$$

Then there exists $\widetilde{a} \in R^n$ such that $f(\widetilde{a}) = 0$ and

$$\widetilde{a} \equiv a \pmod{\mathfrak{p}^m}.$$

Proof. One can reduce from several variables to the one variable version of Hensel's lemma (Lemma 12.2.17): see Exercise 12.11. □

Remark 12.2.19. With essentially the same proof, Hensel's lemma holds more generally for R a complete DVR (without the condition on the residue field) and becomes axiomatically the property of *Henselian* rings.

12.3 Classification via quadratic forms

We now seek to classify quaternion algebras over local fields.

12.3.1. First, suppose F is archimedean. When $F = \mathbb{C}$, the only quaternion algebra over $\mathbb{C}$ up to isomorphism is $B \simeq \mathrm{M}_2(\mathbb{C})$. When $F = \mathbb{R}$, by the theorem of Frobenius (Corollary 3.5.8), there is a unique quaternion division algebra over $\mathbb{R}$.

The classification of quaternion algebras over nonarchimedean local fields is quite analogous to the classification over $\mathbb{R}$, as follows.

Main Theorem 12.3.2. *Let $F \neq \mathbb{C}$ be a local field. Then there is a unique division quaternion algebra B over F up to F-algebra isomorphism.*

We approach the proof of Main Theorem 12.3.2 from two vantage points. In this section, we give a proof using quadratic forms; in the next section, we give another proof by extending the valuation (valid in all characteristics).

To prove this theorem, having dispatched the cases $F = \mathbb{R}, \mathbb{C}$ in 12.3.1 above, from the previous section we may suppose F is a nonarchimedean local field with discrete valuation v, valuation ring R, maximal ideal $\mathfrak{p} = \pi R$ with uniformizer π, and residue field $R/\mathfrak{p} = k$.

12.3.3. Since R is a DVR, all R-lattices M are free (and we only consider those of finite rank): i.e., $M \simeq R^n$ for some $n \in \mathbb{Z}_{\geq 0}$. Given such an R-lattice M, we can reduce modulo $\mathfrak{p}$ to get $M/\mathfrak{p}M \simeq M \otimes_R k \simeq k^n$; conversely, any lift to M of any k-basis of $M/\mathfrak{p}M$ is an R-basis for M, by Nakayama's lemma.

We recall Main Theorem 5.2.5, Corollary 5.2.6, and Main Theorem 5.4.4: isomorphism classes of quaternion algebras over a field F are in natural bijection with nondegenerate ternary quadratic forms up to similarity, and the matrix algebra corresponds to any isotropic form. So to prove Main Theorem 12.3.2, it is equivalent to prove the following statement.

Theorem 12.3.4. *Let $F \neq \mathbb{C}$ be a local field. Then there is a unique anisotropic ternary quadratic form over F up to similarity.*

Rescaling shows equivalently that there is a unique anisotropic ternary quadratic form over a local field $F \neq \mathbb{C}$ of discriminant 1 up to isometry. So our task becomes a hands-on investigation of ternary quadratic forms over F. The theory of quadratic forms over F is linked to that over its residue field k, so we first need to examine isotropy of quadratic forms over a finite field.

Lemma 12.3.5. *A quadratic form $Q \colon V \to k$ over a finite field k with $\dim_k V \geq 3$ is isotropic.*

Proof. The proof is a delightful elementary exercise (Exercise 12.6). □

Recall definitions and notation for quadratic forms over R provided in section 9.7, we embark on a proof in the case where $\operatorname{char} k \neq 2$, beginning with the following lemma.

Lemma 12.3.6. *Suppose* $\operatorname{char} k \neq 2$. *Let* $Q\colon M \to R$ *be a nonsingular quadratic form over* R. *Then the reduction* $Q_k\colon M \otimes_R k \to k$ *of* Q *modulo* $\mathfrak{p}$ *is nonsingular (equivalently, nondegenerate) over* k*; moreover,* Q *is isotropic over* R *if and only if* $Q \bmod \mathfrak{p}$ *is isotropic.*

Proof. For the first statement, by definition we have $\operatorname{disc} Q \in R^\times$, so $\operatorname{disc} Q_k \in k^\times$ by reduction.

For the second, we first prove $(\Rightarrow)$, let $x \in M \smallsetminus \{0\}$ have $Q(x) = 0$. Since Q is homogeneous, we may suppose that $x \notin \mathfrak{p}M$ (divide by powers of π as necessary), so its image in $M \otimes_r k$ is nonzero and thereby shows that Q_k is isotropic. For $(\Leftarrow)$, let $a \in M$ be such that $Q_k(a) = 0 \in k$ and a has nonzero reduction. Choose a basis $M \simeq R^n$ and write $Q(x_1, \dots, x_n) = Q(x_1e_1 + \dots + x_ne_n) \in R[x_1, \dots, x_n]$ in the standard basis as a homogeneous polynomial of degree 2, let T be the associated symmetric bilinear form and $[T] = (T(e_i, e_j))_{i,j}$ the Gram matrix. We are almost ready to apply Hensel's lemma (Lemma 12.2.18), but need to ensure convergence. We observe that

$$\frac{\partial Q}{\partial x_i}(x_1, \dots, x_n) = \sum_{j=1}^{n} T(x_i, x_j)x_j \tag{12.3.7}$$

so the vector of partial derivatives $((\partial Q/\partial x_i)(a))_i = [T]a$ is just the matrix product of the Gram matrix with the vector $a = (a_i)_i$. Working modulo $\mathfrak{p}$, we have $\operatorname{disc} Q_k = 2^{-n}\det[T] \in k^\times$, using that $2 \in k^\times$, so the kernel of $[T] \bmod \mathfrak{p}$ is zero. Since a has nonzero reduction, we conclude that $[T]a$ also has nonzero reduction, which means that $\min_i v((\partial Q/\partial x_i)(a)) = 0$. Therefore the hypotheses of Hensel's lemma are satisfied with $m = 1$, and we conclude there exists a nonzero $\widetilde{a} \in M$ such that $Q(\widetilde{a}) = 0$ and so Q is isotropic. ∎

From Lemma 12.3.6, we obtain the following.

Proposition 12.3.8. *Suppose* $\operatorname{char} k \neq 2$. *Let* $Q\colon M \to R$ *be a nonsingular quadratic form over* R *with* M *free of rank at least* 3. *Then* Q *is isotropic.*

Proof. Combine Lemmas 12.3.5 and 12.3.6. ∎

Considering valuations, we also deduce the following from Lemma 12.3.6.

Lemma 12.3.9. *Suppose* $\operatorname{char} k \neq 2$. *Then* $F^\times/F^{\times 2} \simeq (\mathbb{Z}/2\mathbb{Z})^2$ *and is represented by the classes of* $1, e, \pi, e\pi$ *where* $e \in R^\times$ *is an element which reduces modulo* $\mathfrak{p}$ *to a nonsquare in* k.

We now turn to the proof of our theorem in the case $\operatorname{char} k \neq 2$.

Proof of Theorem 12.3.4 ($\operatorname{char} k \neq 2$). Let $Q \simeq \langle a, -b, -c\rangle$ be an anisotropic ternary quadratic form over F. Then Q is nondegenerate. After rescaling and a change of basis (Exercise 12.8), we may suppose that $a = 1$ and $0 = v(b) \leq v(c)$. If $v(b) = v(c) = 0$ then the quadratic form modulo $\mathfrak{p}$ is nonsingular, so by Lemma 12.3.5 it is isotropic and by Lemma 12.3.6 we conclude Q is isotropic, a contradiction.

We are left with the case $v(b) = 0$ and $v(c) = 1$. By Lemma 12.3.9, we may suppose $b = 1$ or $b = e$ where e is a nonsquare in k. If $b = 1$, then the form is obviously isotropic, so $b = e$. Similarly, $c = \pi$ or $c = e\pi$. In fact, the latter case is similar to the former: dividing by e, we have

$$\langle 1, -e, -e\pi \rangle \sim \langle e^{-1}, -1, -\pi \rangle \simeq \langle -1, e, -\pi \rangle$$

and since $\langle -1, e \rangle \simeq \langle 1, -e \rangle$ (Exercise 12.7), we conclude $Q \sim \langle 1, -e, -\pi \rangle$.

To finish, we show that the form $\langle 1, -e, -\pi \rangle$ is anisotropic. Suppose that $x^2 - ey^2 = \pi z^2$ with $x, y, z \in F^3$ not all zero. By homogeneity, rescaling by a power of π if necessary, we may suppose $x, y, z \in R$ and at least one of $x, y, z \in R^\times$. Reducing modulo $\mathfrak{p}$ we have $x^2 \equiv ey^2 \pmod{\mathfrak{p}}$; since e is a nonsquare in k, we must have $v(x), v(y) \geq 1$. But this implies that $v(z) = 0$ and so $v(\pi z^2) = 1 = v(x^2 - ey^2) \geq 2$, a contradiction. □

Predictably, the proof when $\operatorname{char} k \neq 2$ involving quadratic forms does not generalize in a simple way. However, armed with the above outline and acknowledging these complications, we now pursue the case $\operatorname{char} k = 2$. A key ingredient will understanding certain *binary* quadratic forms, as follows.

Lemma 12.3.10. *There is a unique anisotropic binary quadratic form over k, up to isometry. Moreover, there is a unique (anisotropic) binary quadratic form over R whose reduction modulo $\mathfrak{p}$ is anisotropic, up to isometry (over R).*

When $\operatorname{char} k \neq 2$, this unique class of binary forms over R is $\langle 1, e \rangle$ in the notation above. We will want a similar bit of notation in the case $\operatorname{char} k = 2$.

12.3.11. Recall the issues (6.1.4) with inseparability in characteristic 2. Let $\wp(k) = \{z + z^2 : z \in k\}$ be the **Artin–Schreier group** of k. The polynomial $x^2 + x + a \in k[x]$ is reducible if and only if $a \in \wp(k)$, and since k is finite, $k/\wp(k) \simeq \mathbb{Z}/2\mathbb{Z}$ (Exercise 12.9).

Let $t \in R$ be such that its reduction to k represents the nontrivial class in $k/\wp(k)$.

Proof of Lemma 12.3.10. We begin with the first part of the statement. Let Q be an anisotropic binary quadratic form over k, with T the associated bilinear form. The quadratic form $Q \boxplus \langle -1 \rangle$ is isotropic by Lemma 12.3.5, and Q is anisotropic, so Q represents 1, say $Q(x) = 1$. Extending to a basis, we may rescale the second basis element y so that $T(x, y) = 0, 1$.

- Suppose $T(x, y) = 0$, so $Q = \langle 1, -b \rangle$ for $b \in k^\times$. Since Q is isotropic, $b \notin k^{\times 2}$, so $\operatorname{char} k \neq 2$, the class $b \in k^\times/k^{\times 2}$ is unique, and indeed Q is anisotropic.
- Suppose $T(x, y) = 1$. If $\operatorname{char} k \neq 2$, we may complete the square and reduce to the previous case, so we suppose $\operatorname{char} k = 2$ and $Q(x, y) = x^2 + xy + cy^2$ with $c \in k$. Working now with the Artin–Schreier group, since Q is anisotropic we must have $c + \wp(k) = t + \wp(k) \in k$, giving uniqueness.

For the second statement, if Q is a binary quadratic form over R whose reduction modulo $\mathfrak{p}$ is anisotropic, then we can find a change of basis over k to put Q_k in the

unique isometry class found above; lifting this basis, we may suppose that Q_k is equal to this fixed form. The statement then follows by using Hensel's lemma to lift the identity between any two such lifts, and it is a nice application of Hensel's lemma in two (several) variables: see exercise (Exercise 12.12). □

We now return to the proof of our theorem.

Proof of Theorem 12.3.4 (char $k = 2$). We first claim that the form $[1, 1, t] \boxplus \langle \pi \rangle$ is anisotropic; this follows from a straightforward modification of the argument as in the proof when char $k \neq 2$ above.

We now show this form is the unique one up to similarity. Suppose that Q is a ternary anisotropic form over R. Let $x \in V$ be nonzero; since Q is anisotropic, we may scale x so that $a := Q(x) \in R$. Since $\dim V \geq 3$, there exists nonzero $y' \in V$ such that $T(x, y') = 0$; rescale y' so that $Q(y') \in R$. Let $y := x + y'$. Then $T(x, y) = T(x, x + y') = a$ and $b := Q(y) \in R$, so Q on this basis is $ax^2 + axy + b^2 \simeq [a, a, b]$. We compute that $\operatorname{disc}[a, a, b] = a(a - 4b) \equiv a^2 \pmod{\mathfrak{p}}$, so $\operatorname{disc}[a, a, b] \in R^\times$ and $[a, a, b]$ is nonsingular; completing to a basis with a nonzero element in the orthogonal complement and rescaling, we may suppose without loss of generality that $Q = [a, a, b] \boxplus \langle c \rangle$ with $a, b, c \in R$ and both $v(a) = 0, 1$ and $v(c) = 0, 1$, with $\operatorname{disc} Q = a(a - 4b)c$. This leaves four cases.

- If $v(a) = v(c) = 0$, then reducing modulo $\mathfrak{p}$ we have $\operatorname{disc} Q_k = a^2 c \neq 0$ so Q_k is nondegenerate. By Lemma 12.3.5, Q_k is isotropic. Hensel's lemma (Lemma 12.2.18) applies to Q, showing that Q is isotropic and giving a contradiction—we omit the details.
- Suppose $v(a) = 0$ and $v(c) = 1$. Rescaling by a unit we may suppose $c = \pi$. The reduction $[a, a, b]_k$ modulo $\mathfrak{p}$ is nondegenerate, so if it is isotropic then Q is isotropic, a contradiction. Therefore by Lemma 12.3.10, we may suppose $a = 1$ and $b = t$; this is the desired form.
- Next, consider the case $v(a) = 1$ and $v(c) = 0$; we may suppose $c = 1$. If $v(b) \geq 2$, then $\pi^{-1}[a, a, b] \sim [1, 1, b/\pi]$ so the argument in the previous case applies to show that Q is isotropic. If $v(b) = 1$, then we scale z by π and divide Q by π to reduce to the previous case. Finally, if $v(b) = 0$, we scale y by π and $\pi^{-1}[a, a, b\pi^2] \sim [1, 1, b(\pi^2/a)]$ is again isotropic, a contradiction as in the case $v(b) \geq 2$.
- In the case $v(a) = v(c) = 1$, if $v(b) > 0$ then dividing through by π we reduce to the first case, so we may suppose $v(b) = 0$. Then we multiply by a and replace $x \leftarrow ax$, and $y \leftarrow y/b$ and $z \leftarrow z/\pi$ to work with $[1, a/b, a/b] \boxplus \langle ac/\pi^2 \rangle$, which after interchanging x, y reduces to the previous case $v(a) = 1$, $v(c) = 0$.

Having exhausted the cases and the reader, the result now follows. □

Corollary 12.3.12. *Let F be a nonarchimedean local field with valuation ring R and uniformizer $\pi \in R$. Let B be a quaternion algebra over F.*

If char $k \neq 2$, *then B is a division algebra if and only if*

$$B \simeq \left(\frac{e, \pi}{F}\right), \text{ where } e \in R^\times \text{ is nontrivial in } k^\times/k^{\times 2}$$

and if char F = char $k = 2$*, then* B *is a division algebra if and only if*

$$B \simeq \left[\frac{t, \pi}{F}\right), \text{ where } t \in R \text{ is nontrivial in } k/\wp(k).$$

In Theorem 13.3.11, we rephrase this corollary in terms of the unramified quadratic extension of F.

Remark 12.3.13. In mixed characteristic where char $F = 0$ and char $k = 2$, in the extension $K = F[x]/(x^2 + x + t)$ for t nontrivial in $k/\wp(k)$ we can complete the square to obtain $K = F(\sqrt{e})$ with $e \in F^\times \smallsetminus F^{\times 2}$.

Definition 12.3.14. Let B be a quaternion algebra over F. The **Hasse invariant** of B is defined to be -1 if B is a division algebra and $+1$ if $B \simeq \mathrm{M}_2(F)$.

12.4 Hilbert symbol

Let F be a local field with char $F \neq 2$. We record the splitting behavior of quaternion algebras as follows.

Definition 12.4.1. We define the **Hilbert symbol**

$$(\ ,\ )_F : F^\times \times F^\times \to \{\pm 1\}$$

by the condition that $(a, b)_F = 1$ if and only if the quaternion algebra $\left(\dfrac{a, b}{F}\right) \simeq \mathrm{M}_2(F)$ is split.

The Hilbert symbol is well-defined as a map

$$F^\times/F^{\times 2} \times F^\times/F^{\times 2} \to \{\pm 1\}$$

(Exercise 2.4). By Main Theorem 5.4.4(v), we have $(a, b)_F = 1$ if and only if the **Hilbert equation** $ax^2 + by^2 = 1$ has a solution with $x, y \in F$: this is called **Hilbert's criterion** for the splitting of a quaternion algebra.

Remark 12.4.2. The similarity between the symbols $\left(\dfrac{a, b}{F}\right)$ and $(a, b)_F$ is intentional; but they are not the same, as the former represents an algebra and the latter takes the value ± 1.

In some contexts, the Hilbert symbol $(a, b)_F$ is *defined* to be the isomorphism class of the quaternion algebra $\left(\dfrac{a, b}{F}\right)$ in the Brauer group $\mathrm{Br}(F)$, rather than ± 1 according to whether or not the algebra is split. Conflating these two symbols is not uncommon and in certain contexts it can be quite convenient, but we warn that it can lead to confusion and caution against referring to a quaternion algebra or its isomorphism class as a Hilbert symbol.

Lemma 12.4.3. *Let* $a, b \in F^\times$*. Then the following statements hold:*

(a) $(ac^2, bd^2)_F = (a,b)_F$ *for all* $c,d \in F^\times$.
(b) $(b,a)_F = (a,b)_F$.
(c) $(a,b)_F = (a,-ab)_F = (b,-ab)_F$.
(d) $(1,a)_F = (a,-a)_F = 1$.
(e) *If* $a \neq 1$, *then* $(a, 1-a)_F = 1$.
(f) *If* $\sigma \in \operatorname{Aut}(F)$, *then* $(a,b)_F = (\sigma(a), \sigma(b))_F$.

Proof. Statements (a)–(c) follow from Exercise 2.4. For (d), the Hilbert equation $x^2 + ay^2 = 1$ has the obvious solution $(x,y) = (1,0)$. And $\langle a, -a\rangle$ is isotropic (taking $(x,y) = (1,1)$) so is a hyperbolic plane and represents 1 as in the proof of Main Theorem 5.4.4, or we argue

$$(a,-a)_F = (a,a^2)_F = (a,1)_F = (1,a)_F = 1$$

by Exercise 2.4. For part (e), by Hilbert's criterion $(a,1-a)_F = 1$ since the quadratic equation $ax^2 + (1-a)y^2 = 1$ has the solution $(x,y) = (1,1)$. Finally, part (f): the Hilbert equation $ax^2 + by^2 = 1$ has a solution with $x, y \in F$ if and only if $\sigma(a)x^2 + \sigma(b)y^2 = 1$ has such a solution. □

Remark 12.4.4. Staring at the properties in Lemma 12.4.3 and seeking to axiomatize them, the study of symbols like the Hilbert symbol leads naturally to the definition of $K_2(F)$. In its various formulations, **algebraic K-theory** (K for the German "Klasse", following Grothendieck) seeks to understand certain kinds of functors from rings to abelian groups in a universal sense, encoded in groups $K_n(R)$ for $n \in \mathbb{Z}_{\geq 0}$ and R a commutative ring: see e.g. Karoubi [Kar2010]. For a field F, we have $K_0(F) = \mathbb{Z}$ and $K_1(F) = F^\times$. By a theorem of Matsumoto [Mat69] (see also Milnor [Milno71]), the group $K_2(F)$ is the universal domain for *symbols* over F:

$$K_2(F) := (F^\times \otimes_{\mathbb{Z}} F^\times)/\langle a \otimes (1-a) : a \neq 0,1\rangle.$$

(The tensor product over $\mathbb{Z}$ views $F^\times$ as an abelian group and therefore a $\mathbb{Z}$-module.) The map $a \otimes b \mapsto (a,b)_F$ extends to a map $K_2(F) \to \{\pm 1\}$, a **Steinberg symbol**, a homomorphism from $K_2(F)$ to a multiplicative abelian group. The higher K-groups are related to deeper arithmetic of commutative rings. For an introduction, see Weibel [Weib2013] and Curtis–Reiner [CR87, Chapter 5].

We now turn to be quite explicit about the values of the Hilbert symbol. We begin with the case where F is archimedean. If $F = \mathbb{C}$, then the Hilbert symbol is identically 1. If $F = \mathbb{R}$, then

$$(a,b)_{\mathbb{R}} = \begin{cases} 1, & \text{if } a > 0 \text{ or } b > 0; \\ -1, & \text{if } a < 0 \text{ and } b < 0. \end{cases} \tag{12.4.5}$$

Lemma 12.4.6. *The Hilbert symbol defines a nondegenerate symmetric bimultiplicative pairing*

$$(\,,\,)_F : F^\times/F^{\times 2} \times F^\times/F^{\times 2} \to \{\pm 1\}.$$

By bimultiplicativity, we mean that

$$(a,bc)_F = (a,b)_F(a,c)_F \quad \text{and} \quad (ab,c)_F = (a,c)_F(b,c)_F \tag{12.4.7}$$

for all $a, b, c \in F^\times$ (equivalent, by symmetry).

Keeping in the vibe of this section, we give a proof under the hypothesis that $\operatorname{char} k \neq 2$; for a general proof, see Corollary 13.4.6.

Proof ($\operatorname{char} k \neq 2$*).* This lemma can be read off of the direct computation below (12.4.9), recording what was computed along the way in the proof of Theorem 12.3.4. ∎

12.4.8. Since the Hilbert symbol is well-defined up to squares, the symbol $(a, b)_F$ is determined by the values with $a, b \in \{1, e, \pi, e\pi\}$ where e is a nonsquare in $k^\times$. Let $s = (-1)^{(\#k-1)/2}$, so that $s = 1, -1$ according as -1 is a square in k. Then:

$$\begin{array}{c|cccc} (a,b)_F & 1 & e & \pi & e\pi \\ \hline 1 & 1 & 1 & 1 & 1 \\ e & 1 & 1 & -1 & -1 \\ \pi & 1 & -1 & s & -s \\ e\pi & 1 & -1 & -s & s \end{array} \tag{12.4.9}$$

The computation of this table is requested in Exercise 12.15. For example, if $a = e$, we showed in the proof of Theorem 12.3.4 that $(e, 1)_F = (e, e)_F = 1$, because $ex^2 + y^2 - z^2$ and $ex^2 + ey^2 - z^2$ are isotropic.

In general, writing $a = a_0\pi^{v(a)}$ and $b = b_0\pi^{v(b)}$ we have

$$(a, b)_F = (-1)^{v(a)v(b)(q-1)/2} \left(\frac{a_0}{\mathfrak{p}}\right)^{v(b)} \left(\frac{b_0}{\mathfrak{p}}\right)^{v(a)} \tag{12.4.10}$$

where $q = \#k$ and

$$\left(\frac{c}{\mathfrak{p}}\right) = 0, \pm 1 \equiv c^{(q-1)/2} \pmod{\mathfrak{p}} \tag{12.4.11}$$

is the Legendre symbol: see Exercise 12.16. (Multiplicativity can also be read off of the formula 12.4.10.)

12.4.12. The following easy criteria follow from 12.4.9 (or (12.4.10)):

(a) If $v(ab) = 0$, then $(a, b)_F = 1$.

(b) If $v(a) = 0$ and $v(b) = v(\pi)$, then

$$(a, b)_F = \left(\frac{a}{\mathfrak{p}}\right) = \begin{cases} 1 & \text{if } a \in k^{\times 2}; \\ -1 & \text{if } a \in k^\times \smallsetminus k^{\times 2}. \end{cases}$$

12.4.13. To compute the Hilbert symbol for a local field F with $\operatorname{char} F = 0$ and $\operatorname{char} k = 2$ is significantly more involved. But we can at least compute the Hilbert symbol by hand for $F = \mathbb{Q}_2$.

To begin, the group $\mathbb{Q}_2^\times/\mathbb{Q}_2^{\times 2}$ is generated by $-1, -3, 2$, so representatives are $\{\pm 1, \pm 3, \pm 2, \pm 6\}$. We recall Hilbert's criterion: $(a, b)_F = 1$ if and only if $ax^2 + by^2 = 1$ has a solution with $x, y \in F$.

If $a, b \in \mathbb{Z}$ are odd, then

$$\begin{aligned} ax^2 + by^2 = z^2 &\text{ has a nontrivial solution in } \mathbb{Q}_2 \\ \Leftrightarrow \quad & a \equiv 1 \pmod 4 \text{ or } b \equiv 1 \pmod 4); \end{aligned}$$

by homogeneity and Hensel's lemma, it is enough to check for a solution modulo 4. This deals with all of the symbols with a, b odd: summarizing, we have in this case

$$(a, b)_2 = (-1)^{(a-1)(b-1)/4}. \tag{12.4.14}$$

By the determination above, we see that $(-3, b) = -1$ for $b = \pm 2, \pm 6$ and $(2, 2)_2 = (-1, 2)_2 = 1$ the latter by Hilbert's criterion, as $-1 + 2 = 1$; knowing multiplicativity (Lemma 12.4.6), we have uniquely determined all Hilbert symbols, in particular, for $a \in \mathbb{Z}$ odd we have

$$(a, 2)_2 = (-1)^{(a^2-1)/8}. \tag{12.4.15}$$

It is still useful to compute several of these symbols individually, in the same manner as (12.4.13) (working modulo 8): see Exercise 12.17. We summarize the results here:

$(a, b)_2$	1	−3	−1	3	2	−6	−2	6
1	1	1	1	1	1	1	1	1
−3	1	1	1	1	−1	−1	−1	−1
−1	1	1	−1	−1	1	1	−1	−1
3	1	1	−1	−1	−1	−1	1	1
2	1	−1	1	−1	1	−1	1	−1
−6	1	−1	1	−1	−1	1	−1	1
−2	1	−1	−1	1	1	−1	−1	1
6	1	−1	−1	1	−1	1	1	−1

(12.4.16)

Remark 12.4.17. Analogously, one can define a symbol $[a, b)_F$ for the splitting of quaternion algebras for F a local field with char $F = 2$. This symbol is no longer called the Hilbert symbol, but many properties remain: in particular, there is still an analogue of the **Hilbert equation**, and $\left[\dfrac{a, b}{F}\right)$ is split if and only if $bx^2 + bxy + aby^2 = 1$ has a solution with $x, y \in F$.

Exercises

1. Let p be an odd prime. Show

   (a) Show the equality

   $$(1 - 4x)^{1/2} = 1 - \sum_{n=1}^{\infty} C_n x^n \in \mathbb{Z}[[x]]$$

   of formal series in x with coefficients in $\mathbb{Z}$, where

   $$C_n := \frac{1}{2n-1}\binom{2n}{n} \in \mathbb{Z}_{>0}$$

   are the Catalan numbers. *[Hint: use binomial expansion.]*

(b) Let p be an odd prime. Show that the squaring map is bijective on $1 + p\mathbb{Z}_p$. *[Hint: show that the series expansion in (a) converges in $\mathbb{Z}_p$.]*

2. Recall that a topological space is T_1 if for every pair of distinct points, each point has an open neighborhood not containing the other.

    (a) Show that a topological space X is T_1 if and only if $\{x\}$ is closed for all $x \in X$.
    (b) Let G be a topological group. Show that G is Hausdorff if and only if G is T_1.

3. In this exercise we prove some basic facts about topological groups. Let G be a topological group.

    (a) Let $H \leq G$ be a subgroup. Show that H is open if and only if there exists $h \in H$ and an open neighborhood of h contained in H.
    (b) Show that if $H \leq G$ is an open subgroup, then H is closed.
    (c) Show that a closed subgroup $H \leq G$ of finite index is open.
    (d) Suppose that G is compact. Show that an open subgroup $H \leq G$ is of finite index, and that every open subgroup contains an open *normal* subgroup.

▹ 4. Let G be a topological group. Let $U \ni 1$ be an open neighborhood of 1.

    (a) Show that there exists an open neighborhood $V \subseteq U$ of $1 \in V$ such that $V^2 = V \cdot V \subseteq U$. *[Hint: Multiplication is continuous.]*
    (b) Similarly, show that there exists an open neighborhood $V \subseteq U$ of $1 \in V$ such that $V^{-1}V \subseteq U$.

5. Let G be a topological group and let $H \leq G$ be a closed subgroup. Equip G/H with the quotient topology. Show that G/H is Hausdorff. *[Hint: Use Exercise 12.4(b).]*

▹ 6. Let k be a finite field and let $Q\colon V \to k$ be a ternary quadratic form. Show that Q is isotropic. *[Hint: Reduce to the case of finding a solution to $y^2 = f(x)$ where f is a polynomial of degree 2. If $\#k$ is odd, count squares and the number of distinct values taken by $f(x)$ in k. Second approach: reduce to the case where $\#k$ is odd, and show that $x^2 + y^2$ represents a nonsquare, since the squares cannot be closed under addition!]* [This repeats Exercise 5.5!]

▹ 7. Let k be a finite field with $\operatorname{char} k \neq 2$ and let $e \in k^\times$. Show directly that there is an *isometry* $\langle -1, e\rangle \simeq \langle 1, -e\rangle$.

▹ 8. Let R be a DVR with field of fractions F, let $a, b, c \in F$ be nonzero and let $Q = \langle a, -b, -c\rangle$. Show that Q is similar over F to $\langle 1, -b', -c'\rangle$ with $0 = v(b') \leq v(c')$. *[Hint: first get $v(a), v(b), v(c) \in \{0, 1\}$.]*

▹ 9. Let k be a finite field with even cardinality. Show that $\#k/\wp(k) = 2$, where $\wp(k)$ is the Artin-Schreier group.

10. By Theorem 12.2.15, a complete archimedean local field is isomorphic to $\mathbb{R}$ or $\mathbb{C}$. Extend this classification to division algebras as follows.

    The notion of *absolute value* (Definition 12.2.2) extends to a division algebra without modification, as does the notion of *archimedean* and *nonarchimedean*.

    (a) Show that $\mathbb{H}$ has an absolute value $|\alpha| = \sqrt{\operatorname{nrd}(\alpha)}$ for $\alpha \in \mathbb{H}$.
    (b) Let D be a division algebra equipped with an absolute value $|\ |$. Show that if $|\ |$ is archimedean, then $\operatorname{char} D = 0$ and if the restriction of $|\ |$ to its center $Z(D)$ is archimedean.
    (c) Show that every division algebra complete with respect to an archimedean absolute value is isomorphic to $\mathbb{R}$, $\mathbb{C}$, or $\mathbb{H}$ and with the absolute value equivalent to the absolute value $|\alpha| = \sqrt{\operatorname{nrd}\alpha}$ in each case. *[Hint: recall Theorem 3.5.1.]*

▹ 11. Prove Lemma 12.2.18 using Lemma 12.2.17. *[Hint: let j be the index that achieves the minimal valuation among partial derivatives, and consider the restriction $f(a_1, \dots, a_{j-1}, x, a_{j+1}, \dots, a_n) \in R[x]$ to one variable.]*

▹ 12. In this exercise, we consider an extension of Hensel's lemma to several polynomials (in several variables). Let F be a nonarchimedean local field with valuation v and valuation ring R.

    (a) Let $f_1, \dots, f_n \in R[x_1, \dots, x_n]$ be polynomials with $n \geq 1$. Let $a \in R^n$ have $m := \min_i v(f_i(a))$. Let
    $$\mathrm{D}_f := \left(\frac{\partial f_i}{\partial x_j}(a)\right)_{i,j} \in M_n(R).$$
    Suppose $m > 2v(\det \mathrm{D}_f(a))$. Show that there exists $\widetilde{a} \in R^n$ such that $f_i(a) = 0$ for all $i = 1, \dots, n$ and $\widetilde{a} \equiv a \pmod{\mathfrak{p}^m}$. *[Hint: by Taylor expansion, write $f(a+\pi^t x) = f(a) + \mathrm{D}_f(a)p^t x + p^{2t} r(x)$ in vector notation with $t \geq m$, and iteratively solve the system in a manner analogous to Newton's method. See Conrad [Con] for a complete development.]*

    (b) Show that (a) also holds more generally if the number of polynomials r has $r \leq n$ and there exists an $r \times r$ matrix minor of $\mathrm{D}_f(a)$ whose determinant has valuation $< m/2$. *[Hint: see Exercise 12.11.]*

    (c) Finish the proof of Lemma 12.3.10 as follows. Let Q, Q' be two binary quadratic forms over R such that their reductions Q_k, Q'_k are anisotropic. Show that $Q \simeq Q'$. *[Hint: reduce to the case where $Q_k = Q'_k$. If $Q_k \simeq \langle 1, -e \rangle$, rescale the basis using a plain vanilla flavor of Hensel's lemma. If $Q_k \simeq [1,1,t]$, reduce to showing that $Q \simeq [1,1,t]$. Consider a general change of variables in $\mathrm{GL}_2(R)$ from $[1,1,t]$ that reduces to the identity modulo $\mathfrak{p}$ and apply the deluxe version of Hensel's lemma in part (b) to the resulting system of three equations in four unknowns: there is a minor with determinant $1-4t \in R^\times$.]*

13. Let $F \neq \mathbb{C}$ be a local field and let Q be a nondegenerate ternary quadratic form over F. Let $K \supseteq F$ be a quadratic field extension. Show that Q is isotropic over K.

14. Give another proof of Lemma 12.4.6 that the local Hilbert symbol is bimultiplicative using Example 8.2.2 and the Brauer group (section 8.3).

▹ 15. Show that the table of Hilbert symbols (12.4.9) is correct.

▹ 16. One can package 12.4.8 together with multiplying by squares to prove the following more general criterion. Let F be a nonarchimedean local field with uniformizer π, valuation v with $v(\pi) = 1$, and residue field k. Let $q = \#k$ and suppose q is odd.
Show that for $a, b \in F^\times$, if we write $a = a_0\pi^{v(a)}$ and $b = b_0\pi^{v(b)}$, then

$$(a, b)_F = (-1)^{v(a)v(b)(q-1)/2}\left(\frac{a_0}{\pi}\right)^{v(b)}\left(\frac{b_0}{\pi}\right)^{v(a)}.$$

▹ 17. Show that the table of Hilbert symbols (12.4.16) is correct by considering the equation $ax^2 + by^2 \equiv 1 \pmod 8$.

18. Prove a descent for the Hilbert symbol, as follows. Let K be a finite extension of the local field F with $\operatorname{char} F \neq 2$ and let $a, b \in F^\times$. Show that $(a, b)_K = (a, \mathrm{Nm}_{K|F}(b))_F = (\mathrm{Nm}_{K|F}(a), b)_F$.

This chapter is licensed under the terms of the Creative Commons Attribution-NonCommercial 4.0 International License (http://creativecommons.org/licenses/by-nc/4.0/), which permits any noncommercial use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license and indicate if changes were made.

The images or other third party material in this chapter are included in the chapter's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the chapter's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.

Chapter 13

Quaternion algebras over local fields

In this chapter, we approach the classification of quaternion algebras over local fields in a second way, using valuations.

13.1 Extending the valuation

Recall (section 12.1) the valuation $v = v_p$ on $\mathbb{Q}_p$, measuring divisibility by p. We have

$$\begin{aligned} \mathbb{Z}_p &= \{x \in \mathbb{Q}_p : v(x) \geq 0\}, \text{ and} \\ p\mathbb{Z}_p &= \{x \in \mathbb{Q}_p : v(x) > 0\}; \end{aligned} \tag{13.1.1}$$

Indeed, these can profitably be taken as their definition.

For any finite extension $K \supseteq \mathbb{Q}_p$ of fields, there is a unique valuation w on K such that $w|_{\mathbb{Q}_p} = v$ (so w **extends** v), defined by

$$w(x) := \frac{v(\mathrm{Nm}_{K|\mathbb{Q}_p}(x))}{[K : \mathbb{Q}_p]}. \tag{13.1.2}$$

The integral closure of $\mathbb{Z}_p$ in K is the valuation ring $\{x \in K : w(x) \geq 0\} \supseteq \mathbb{Z}_p$, and its unique maximal ideal is $\{x \in K : w(x) > 0\}$, as in (13.1.1).

For example, there is a unique unramified quadratic extension K of $\mathbb{Q}_p$: we have $K = \mathbb{Q}_p(\sqrt{e})$, where $e = -3$ for $p = 2$ and otherwise $e \in \mathbb{Z}$ is a quadratic nonresidue modulo p for p odd. It is common to write $K = \mathbb{Q}_{p^2}$ for this field and $\mathbb{Z}_{p^2}$ for its valuation ring, since the residue field of K is $\mathbb{F}_{p^2}$.

In a completely parallel fashion, let B be a *division* quaternion algebra over $\mathbb{Q}_p$. Then there is again a unique valuation w extending v, defined by

$$\begin{aligned} w \colon B &\to \mathbb{R} \cup \{\infty\} \\ \alpha &\mapsto \frac{v(\mathrm{nrd}(\alpha))}{2}. \end{aligned} \tag{13.1.3}$$

The valuation ring

$$O := \{\alpha \in B : w(\alpha) \geq 0\} \tag{13.1.4}$$

© The Author(s) 2021

J. Voight, *Quaternion Algebras*, Graduate Texts in Mathematics 288,
https://doi.org/10.1007/978-3-030-56694-4_13

is the unique (!) maximal R-order in B, consisting of all elements of B that are integral over $\mathbb{Z}_p$. The set

$$P := \{\alpha \in B : w(\alpha) > 0\} \tag{13.1.5}$$

is the unique maximal two-sided (bilateral) ideal of O.

Using the unique extension of the valuation, we obtain the following main result of this chapter (a special case of Theorem 13.3.11).

Theorem 13.1.6. *Let $q := p^2$. Then the following statements hold.*

(a) *There is a unique division quaternion algebra B over $\mathbb{Q}_p$, up to isomorphism given by $B \simeq \left(\dfrac{\mathbb{Q}_q, p}{\mathbb{Q}_p}\right)$.*
(b) *The valuation ring of B is $O \simeq \mathbb{Z}_q \oplus \mathbb{Z}_q j$.*
(c) *The maximal ideal $P = Oj$ has $P^2 = pO$ and $O/P \simeq \mathbb{Z}_q/p\mathbb{Z}_q \simeq \mathbb{F}_q$.*

The method of proof used in this classification can also be used to classify central division algebras over local fields in much the same manner.

13.2 Valuations

To begin, we briefly review extensions of valuations; for further reading, see the references given in section 12.2.

Let R be a *complete* DVR with valuation $v\colon R \to \mathbb{Z}_{\geq 0} \cup \{\infty\}$, field of fractions F, maximal ideal $\mathfrak{p}$ generated by a uniformizer π (with $v(\pi) = 1$), and residue field $k := R/\mathfrak{p}$. Then R is an integrally closed PID (every ideal is a power of the maximal ideal $\mathfrak{p}$), and $R = \{x \in F : v(x) \geq 0\}$. Let $|\,|_v$ be an absolute value attached to v, as in (12.2.11).

Let $K \supseteq F$ be a finite separable extension of degree $n := [K : F]$. Then in fact K is also a nonarchimedean local field; more precisely, we have the following lemma.

Lemma 13.2.1. *There exists a unique valuation w on K such that $w|_F = v$, defined by*

$$w(x) := \frac{v(\mathrm{Nm}_{K|F}(x))}{[K : F]}. \tag{13.2.2}$$

The integral closure of R in K is the valuation ring

$$S := \{x \in K : w(x) \geq 0\}.$$

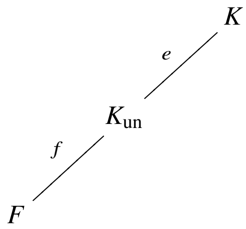
When $w|_F = v$, we say that w **extends** v.

Proof. See e.g. Neukrich [Neu99, Chapter II, Theorem (4.8)], Cassels [Cas86, Chapter 7, Theorem 1.1], or Serre [Ser79, Chapter II, §2, Proposition 3]. □

By the same token using (13.2.2), there exists a unique absolute value $|\,|_w$ on K which restricts to $|\,|_v$ on F; we pass freely between these two formulations.

13.2.3. We say $K \supseteq F$ is **unramified** if a uniformizer π for F is also a uniformizer for K. We say $K \supseteq F$ is **totally ramified** if a uniformizer π_K has the property that π_K^n is a uniformizer for F.

In general, there is a (unique) maximal unramified subextension $K_{\text{un}} \subseteq K$, and the extension $K \supseteq K_{\text{un}}$ is totally ramified.



We say that $e = [K : K_{\text{un}}]$ is the **ramification degree** and $f = [K_{\text{un}} : F]$ the **inertial degree**, and the fundamental equality

$$n = [K : F] = ef \tag{13.2.4}$$

holds.

13.2.5. Suppose that F is a local field (equivalently, the residue field k is a finite field). Then for all $f \in \mathbb{Z}_{\geq 1}$, there is a unique unramified extension of F of degree f and such a field corresponds to the unique extension of the residue field k of degree f. In an unramified extension $K \supseteq F$ of degree $[K : F] = f$, we have $\operatorname{Nm}_{K|F}(K^\times) = R^\times \pi^{f\mathbb{Z}}$, so $b \in \operatorname{Nm}_{K|F}(K^\times)$ if and only if $f \mid v(b)$.

If $\operatorname{char} k \neq 2$, then by Hensel's lemma, the unramified extension of degree 2 is given by adjoining a square root of an element of R which reduces to the unique nontrivial class in $k^\times/k^{\times 2}$; if $\operatorname{char} k = 2$, then the unramified extension of degree 2 is given by adjoining a root of the polynomial $x^2 + x + t$ where $t \in R$ reduces to an element which is nontrivial in the Artin-Schreier group $k/\wp(k)$ (recalling 12.3.11).

Before proceeding further, we describe local fields by their defining polynomials—we will need this later in the study of norms and strong approximation.

Lemma 13.2.6. (Krasner's lemma). *Let $K \supseteq F$ be a finite, Galois extension with absolute value $|\ |_w$. Let $\alpha, \beta \in K$, and suppose that for all $\sigma \in \operatorname{Gal}(K \mid F)$ with $\sigma(\alpha) \neq \alpha$, we have*

$$|\alpha - \beta|_w < |\alpha - \sigma(\alpha)|_w. \tag{13.2.7}$$

Then $F(\alpha) \subseteq F(\beta)$.

Intuitively, we can think of Krasner's lemma as telling us when β is closer to α than any of its conjugates, then $F(\beta)$ contains α. It is for this reason that we state the lemma in terms of absolute values (instead of valuations).

Proof. Let $\sigma \in \operatorname{Gal}(K \mid F(\beta))$ have $\sigma(\alpha) \neq \alpha$. Then by the ultrametric inequality,

$$\begin{aligned} |\sigma(\alpha) - \alpha|_w &= |\sigma(\alpha) - \beta + \beta - \alpha|_w \leq \max(|\sigma(\alpha) - \beta|_w, |\beta - \alpha|_w) \\ &= \max(|\sigma(\alpha - \beta)|_w, |\beta - \alpha|_w) = |\alpha - \beta|_w, \end{aligned} \tag{13.2.8}$$

the final equality a consequence of (13.2.2) and the fact that Galois conjugates have the same norm. This contradicts the existence of σ, so $\sigma(\alpha) = \alpha$ for all $\sigma \in \mathrm{Gal}(K \mid F(\beta))$. By Galois theory, we conclude that $F(\alpha) \subseteq F(\beta)$. ∎

Corollary 13.2.9. *Let $f(x) = x^n + a_{n-1}x^{n-1} + \cdots + a_0 \in F[x]$ be a separable, monic polynomial. Then there exists $\delta > 0$ such that whenever $g(x) = x^n + b_{n-1}x^{n-1} + \cdots + b_0 \in F[x]$ has $|a_i - b_i|_v < \delta$, then*

$$F[x]/(f(x)) \simeq F[x]/(g(x)).$$

In particular, if $f(x)$ is irreducible then $g(x)$ is irreducible.

Proof. Since $f(x)$ is separable, its discriminant $\mathrm{disc}(f)$ is nonzero. The discriminant is a polynomial function in the coefficients, so by continuity (multivariate Taylor expansion), there exists $\delta_1 > 0$ such that if $g(x) = x^n + \cdots + b_0 \in F[x]$ has $|a_i - b_i|_v < \delta_1$ for all i, then $|\mathrm{disc}(g) - \mathrm{disc}(f)|_v < |\mathrm{disc}(f)|_v$; by the ultrametric inequality, we conclude that for such $g(x)$ we have $|\mathrm{disc}(g)|_v = |\mathrm{disc}(f)|_v$ so in particular $\mathrm{disc}(g) \neq 0$.

Let $g(x)$ be as in the previous paragraph; then $g(x)$ is separable. We first consider the case where $f(x)$ is irreducible. Let $K \supseteq F$ be a splitting field for the polynomials $f(x) = \prod_{i=1}^n (x - \alpha_i) \in K[x]$ and $g(x) = \prod_{i=1}^n (x - \beta_i)$. Let $|\ |_w$ on K extend $|\ |_v$. Let

$$\epsilon := \min_{i \neq j} |\alpha_i - \alpha_j|_w. \tag{13.2.10}$$

Finally, let

$$\rho(g) = \rho(b_0, \dots, b_{n-1}) := \prod_{i=1}^n g(\alpha_i) = \prod_{i,j=1}^n (\alpha_i - \beta_j). \tag{13.2.11}$$

The map $g \mapsto \rho(g)$ is again a polynomial in the coefficients $b_0, \dots, b_{n-1}$ (indeed, it is a polynomial resultant). Therefore there exists $\delta > 0$, with $\delta < \delta_1$, such that if $|a_i - b_i|_v < \delta$, then $|\rho(g)|_w < \epsilon^{n^2}$. Therefore in (13.2.11), there exists i, j such that $|\alpha_i - \beta_j|_w < \epsilon$. Together with (13.2.10), we have

$$|\alpha_i - \beta_j|_w < \epsilon \leq |\alpha_i - \alpha_k|_w$$

for all $k \neq i$. By Krasner's lemma (Lemma 13.2.6), we conclude that $F(\alpha_i) \subseteq F(\beta_j)$. Since $f(\alpha_i) = g(\beta_j) = 0$ and f is irreducible, we have

$$[F(\alpha_i) : F] = n \leq [F(\beta_j) : F] \leq n$$

so in fact $[F(\beta_j) : F] = n$ and $F(\alpha_i) = F(\beta_j)$. Finally,

$$F[x]/(f(x)) \simeq F(\alpha_i) = F(\beta_j) \simeq F[x]/(g(x))$$

as desired.

The case when $f(x) = f_1(x) \cdots f_r(x)$ is reducible follows by repeating the above argument on each factor, and finishing using the continuity of multiplication among the coefficients: the details are requested in Exercise 13.16. ∎

13.3 Classification via extensions of valuations

We now seek to generalize this setup to the noncommutative case; we retain the notation from the previous section. Let D be a central (simple) *division* algebra over F with $\dim_F D = [D : F] = n^2$. We extend the valuation v to a map

$$\begin{aligned} w : D &\to \mathbb{R} \cup \{\infty\} \\ \alpha &\mapsto \frac{v(\mathrm{Nm}_{D|F}(\alpha))}{[D:F]} = \frac{v(\mathrm{nrd}(\alpha))}{n}, \end{aligned} \tag{13.3.1}$$

where the equality follows from the fact that $\mathrm{Nm}_{D|F}(\alpha) = \mathrm{nrd}(\alpha)^n$ (see section 7.8).

Lemma 13.3.2. *The map w is the unique valuation on D extending v, i.e., the following hold:*

(i) $w(\alpha) = \infty$ *if and only if* $\alpha = 0$.
(ii) $w(\alpha\beta) = w(\alpha) + w(\beta) = w(\beta\alpha)$ *for all* $\alpha, \beta \in D$.
(iii) $w(\alpha + \beta) \geq \min(w(\alpha), w(\beta))$ *for all* $\alpha, \beta \in D$.
(iv) $w(D^\times)$ *is discrete in* $\mathbb{R}$.

Proof. Since D is a division ring, statement (i) is immediate. Statement (ii) follows from the multiplicativity of nrd and v. To prove (iii), we may suppose $\beta \neq 0$ and so $\beta \in D^\times$. We have

$$w(\alpha + \beta) = w((\alpha\beta^{-1} + 1)\beta) = w(\alpha\beta^{-1} + 1) + w(\beta).$$

But the restriction of w to $F(\alpha\beta^{-1})$ is a discrete valuation, thus $w(\alpha\beta^{-1} + 1) \geq \min(w(\alpha\beta^{-1}), w(1))$ and by (ii) $w(\alpha + \beta) \geq \min(w(\alpha), w(\beta))$, as desired. Finally, (iv) holds since $w(D^\times) \subseteq v(F^\times)/n$ and the latter is discrete. The valuation is unique because it is unique whenever it is restricted to a subfield. □

13.3.3. From Lemma 13.3.2, we say that w is a **discrete valuation** on D since it satisfies the same axioms as for a field. It follows from Lemma 13.3.2 that the set

$$\mathcal{O} := \{\alpha \in D : w(\alpha) \geq 0\}$$

is a ring, called the **valuation ring** of D.

Proposition 13.3.4. *The ring $\mathcal{O}$ is the unique maximal R-order in D, consisting of all elements of D that are integral over R.*

Proof. First, we prove that

$$\mathcal{O} = \{\alpha \in D : \alpha \text{ is integral over } R\}. \tag{13.3.5}$$

We first show the inclusion ($\supseteq$) of (13.3.5), and suppose $\alpha \in D$ is integral over R. Since R is integrally closed, by Lemma 10.3.5 the coefficients of the minimal polynomial $f(x) \in F[x]$ of α belong to R. Since D is a division ring, $f(x)$ is irreducible and hence the reduced characteristic polynomial $g(x)$ is a power of $f(x)$ and thus has

coefficients in R. Up to sign, the constant coefficient of $g(x)$ is $\operatorname{nrd}(\alpha)$, so $w(\alpha) = v(\operatorname{nrd}(\alpha)) \geq 0$, hence $\alpha \in O$.

Next we prove ($\subseteq$) in (13.3.5). Suppose $\alpha \in O$, so that $w(\alpha) \geq 0$, and let $K = F(\alpha)$. Let $f(x) \in F[x]$ be the minimal polynomial of α. We want to conclude that $f(x) \in R[x]$ knowing that $w(\alpha) \geq 0$. But the restriction of w to K is the unique extension of v to K, and this is a statement about the extension $K \supseteq F$ of fields and therefore follows from the commutative case, Lemma 13.2.1.

We can now prove that O is an R-order. Scaling an element of $D^\times$ by an appropriate power of π gives it positive valuation, so $OF = D$. To conclude, we must show that O is finitely generated as an R-module. Recall that D is a central division algebra over F, hence a separable F-algebra, so we may apply Lemma 10.3.7: every $\alpha \in O$ is integral over R and O is a ring, and the lemma implies that O is an R-order.

Finally, it follows immediately that O is a maximal R-order: by Corollary 10.3.3, every element of an R-order is integral over R, and O contains all such elements. □

Remark 13.3.6. For a quaternion division algebra D, we can argue more directly in the proof of Proposition 13.3.4 using the reduced norm: see Exercise 13.4.

13.3.7. It follows from Proposition 13.3.4 that O is a finitely generated R-submodule of D. But R is a PID so in fact O is free of rank $[D : F]$ as an R-module. We have

$$O^\times = \{\alpha \in D : w(\alpha) = 0\} \tag{13.3.8}$$

since $w(\alpha^{-1}) = -w(\alpha)$, and in particular $\alpha \in O^\times$ if and only if $\operatorname{nrd}(\alpha) \in R^\times$. Consequently,

$$P := \{\alpha \in D : w(\alpha) > 0\} = O \smallsetminus O^\times \tag{13.3.9}$$

is the unique maximal two-sided (bilateral) ideal of O, as well as the unique left or right ideal of O. Therefore O is a **noncommutative local ring**, a noncommutative ring with a unique maximal left (equivalently, right) ideal.

13.3.10. Let $\beta \in P$ have minimal (positive) valuation $w(\beta) > 0$. Then for all $0 \neq \alpha \in P$ we have $w(\alpha\beta^{-1}) = w(\alpha) - w(\beta) \geq 0$ so $\alpha\beta^{-1} \in O$ and $\alpha \in O\beta$. Arguing on the other side, we have also $\alpha \in \beta O$. Thus $P = O\beta = \beta O = O\beta O$.

Arguing in the same way, we see that every one-sided ideal of O is in fact two-sided, and every two-sided ideal of O is principally generated by any element with minimal valuation hence of the form P^r for some $r \in \mathbb{Z}_{\geq 0}$.

We are now prepared to give the second proof of the main result in this chapter (Main Theorem 12.3.2). We now add the hypothesis that F is a local field, so that k is a finite field.

Theorem 13.3.11. *Let F be a nonarchimedean local field. Then the following statements hold.*

(a) *There is a unique division quaternion algebra B over F, up to F-algebra isomorphism given by*

$$B \simeq \left(\frac{K, \pi}{F}\right),$$

where K is the unique quadratic unramified (separable) extension of F.

(b) *Let B be as in* (a). *Then the valuation ring of B is $\mathcal{O} \simeq S \oplus Sj$, where S is the integral closure of R in K. Moreover, the ideal $P = \mathcal{O}j$ is the unique maximal ideal; we have $P^2 = \pi\mathcal{O}$, and $\mathcal{O}/P \supseteq R/\mathfrak{p}$ is a quadratic extension of finite fields.*

Proof. We begin with existence in part (a), and existence: we prove that $B = (K, \pi \mid F)$ is a division algebra. We recall that B is a division ring if and only if $\pi \notin \mathrm{Nm}_{K|F}(K^\times)$ by Main Theorem 5.4.4 and Theorem 6.4.11. Since $K \supseteq F$ is unramified, we have $\mathrm{Nm}_{K|F}(K^\times) = R^\times\pi^{2\mathbb{Z}}$ by 13.2.5. Putting these together gives the result.

Continuing with (a), we now show uniqueness. Let B be a division quaternion algebra over F. We refer to 13.3.10, and let $P = \mathcal{O}\beta$. Then $w(\beta) \in \frac{1}{2}\mathbb{Z}_{>0}$, so

$$w(\beta) \le w(\pi) = v(\pi) = 1 \le 2w(\beta) = w(\beta^2); \tag{13.3.12}$$

we conclude that $\beta\mathcal{O} = P \supseteq \pi\mathcal{O} \supseteq P^2 = \beta^2\mathcal{O}$. The map $\alpha \mapsto \alpha\beta$ yields an isomorphism $\mathcal{O}/P \xrightarrow{\sim} P/P^2$ of k-vector spaces, so

$$4 = \dim_k(\mathcal{O}/\pi\mathcal{O}) \le \dim_k(\mathcal{O}/P^2) = \dim_k(\mathcal{O}/P) + \dim_k(P/P^2) = 2\dim_k(\mathcal{O}/P) \tag{13.3.13}$$

and thus $\dim_k(\mathcal{O}/P) \ge 2$, with equality if and only if $\pi\mathcal{O} = P^2$.

As in (13.3.9), we have $\mathcal{O}\smallsetminus P = \mathcal{O}^\times$, so the ring $\mathcal{O}/P$ is a division algebra over k and hence a finite division ring. By Wedderburn's little theorem (Exercise 7.30), we conclude that $\mathcal{O}/P$ is a finite field! So there exists $i \in \mathcal{O}$ such that its reduction generates $\mathcal{O}/P$ as a finite extension of k. But i satisfies its reduced characteristic polynomial, a monic polynomial of degree 2 with coefficients in R, so its reduction satisfies a polynomial of degree 2 with coefficients in k. Since i is a generator, we conclude $[\mathcal{O}/P : k] \le 2$. Together with the conclusion of the previous paragraph, we conclude that $[\mathcal{O}/P : k] = \dim_k(\mathcal{O}/P) = 2$, in other words $\mathcal{O}/P$ is a (separable) quadratic field extension of k. It then follows from 13.2.5 that $K := F(i)$ is the unique, unramified (separable) quadratic extension of F. Therefore equality holds in (13.3.13) and $P^2 = \pi\mathcal{O}$. Since $\beta^2\mathcal{O} = P^2 = \pi\mathcal{O}$, we have $w(\beta) = 1/2$.

By Exercise 6.2 or 7.27, there exists $b \in F^\times$ such that $B \simeq (K, b \mid F)$. Recalling the first paragraph above, since B is a division algebra, we have $b \notin \mathrm{Nm}_{K|F}(K^\times) = R^\times\pi^{2\mathbb{Z}}$. Applying Exercise 6.4, we may multiply b by a norm from $K^\times$, so we may suppose $b = \pi$, and therefore $B \simeq (K, \pi \mid F)$. This concludes the proof of (a).

We turn now to (b), with $B = (K, \pi \mid F) = K + Kj$ with $j^2 = \pi$. Let $\alpha = u + vj \in B$ with $u, v \in K$. Then $\mathrm{nrd}(\alpha) = \mathrm{Nm}_{K|F}(u) - \pi\,\mathrm{Nm}_{K|F}(v) = x - \pi y$ with $x, y \in F$ and $v(x)$ even and $v(\pi y)$ odd (as norms from K). By the ultrametric inequality, we have $w(\alpha) = v(\mathrm{nrd}(\alpha)) \ge 0$ if and only if $v(x), v(y) \ge 0$ if and only if $u, v \in S$ (as S is the valuation ring of K). Therefore $\mathcal{O} = S + Sj$. Since $j^2 = \pi$, we have $w(j) = 1/2$, so $j\mathcal{O} = P$. The remaining statements were proven in the course of proving (a). ◻

Corollary 13.3.14. *Let $F \neq \mathbb{C}$ be a local field. Let K be the unramified quadratic extension of F, with $\langle\sigma\rangle = \mathrm{Gal}(K \mid F)$. Then the F-subalgebra*

$$B = \left\{ \begin{pmatrix} u & \pi v \\ \sigma(v) & \sigma(u) \end{pmatrix} : u, v \in K \right\} \subset \mathrm{M}_2(K)$$

is the unique division quaternion algebra over F (up to isomorphism).

Proof. Using Theorem 13.3.11(a), we split B over K as in 2.3.4. (We may also put π below the diagonal as in 2.3.12.) $\square$

13.4 Consequences

We now observe a few consequences of Theorem 13.3.11.

Corollary 13.4.1. *Let F be a nonarchimedean local field with valuation v, let K be a separable, unramified quadratic F-algebra, and let $B = (K, b \mid F)$ with $b \in F^\times$. If $v(b) = 0$, then $B \simeq \mathrm{M}_2(F)$.*

Proof. Either $K \simeq F \times F$ or $K \supseteq F$ is the unique unramified quadratic field extension. In the first case, K has a zerodivisor so $B \simeq \mathrm{M}_2(F)$. In the second case, we conclude as in the first paragraph of the proof of Theorem 13.3.11, since $b \in R^\times \leq \mathrm{Nm}_{K|F}(K^\times)$. $\square$

13.4.2. Let B be a division quaternion algebra over F. In analogy with the case of field extensions (13.2.4), we define the **ramification index** of B over F as $e(B|F) = 2$ since $P^2 = \pi O$, and the **inertial degree** of B over F as $f(B|F) = 2$ since B contains the unramified quadratic extension K of F, and note the equality

$$e(B|F)f(B|F) = 4 = [B : F],$$

as in the commutative case. (Viewed in this way, B is obtained from first an unramified extension and then a "noncommutative" ramified extension.)

Remark 13.4.3. Theorem 13.3.11, the fundamental result describing division quaternion algebras over a local field, is a special case of a more general result as follows. Let R be a complete DVR with maximal ideal $\mathfrak{p} = \pi R$ and $F := \mathrm{Frac}(R)$.

Let D be a (finite-dimensional) division algebra over F, and let $O \subsetneq D$ be the valuation ring and $P \subset O$ the maximal ideal. Then $P^e = \mathfrak{p}O$ for some $e \geq 1$, called the **ramification index**; the quotient O/P is a division algebra over the field $k = R/\mathfrak{p}$, and we let the **inertial degree** be $f = \dim_k(O/P)$. Then $ef = \dim_F D = n^2$; moreover, if k is finite (F is a local field), then $e = f = n$. For a proof, see Exercise 13.10; or consult Reiner [Rei2003, Theorems 12.8, 13.3, 14.3]. However, the uniqueness of D up to F-algebra isomorphism no longer holds. If F is a local field, then the possibilities for D are classified up to isomorphism by a local invariant $\mathrm{inv}\, D \in (\frac{1}{n}\mathbb{Z})/\mathbb{Z} \simeq \mathbb{Z}/n\mathbb{Z}$. These patch together to give a global result: see Remark 14.6.10.

This classification can be further extended to an arbitrary central simple algebra $B \simeq \mathrm{M}_n(D)$ over F: see Reiner [Rei2003, §17–18].

Splitting of local division quaternion algebras over extension fields is given by the following simple criterion.

Proposition 13.4.4. *Let B be a division quaternion algebra over a local field F, and let L be a separable field extension of F of finite degree. Then L is a splitting field for B if and only if $[L : F]$ is even.*

Proof. If F is archimedean, then either $F = \mathbb{C}$ and there is no such L, or $F = \mathbb{R}$ and $B = \mathbb{H}$ and $L = \mathbb{C}$, and the result holds. So suppose F is nonarchimedean. We have $B \simeq (K, \pi \mid F)$ where K is the unramified quadratic extension of F. Let e, f be the ramification index and inertial degree of L, respectively. Then $[L : F] = n = ef$, and n is even if and only if e is even or f is even. But f is even if and only if L contains an unramified quadratic subextension, necessarily isomorphic to K; but then K splits B so L splits B.

Having established the claim when f is even, suppose that f is odd. Then L is linearly disjoint from K and $K \otimes_F L = KL$ is the unramified quadratic extension of L. Therefore $B \otimes_F L \simeq (KL, \pi \mid L)$. Let R_L be the valuation ring of L and let π_L be a uniformizer for L. Then $\mathrm{Nm}_{KL/L}(KL^\times) = R_L^\times \pi_L^{2\mathbb{Z}}$. We have $\pi = u\pi_L^e$ for some $u \in R_L^\times$. Putting these together, we see that $B \otimes_F L$ is split if and only if π is a norm from KL if and only if e is even. □

As a consequence, B contains every separable quadratic extension of F.

Corollary 13.4.5. *If B is a division quaternion algebra over a local field F and $K \supseteq F$ is a separable quadratic field extension, then $K \hookrightarrow B$.*

Proof. Combine Proposition 13.4.4 with Lemmas 5.4.7 and 6.4.12. □

We repeat now Lemma 12.4.6, giving a proof that works without restriction on characteristic.

Corollary 13.4.6. *If* char $F \neq 2$, *then the Hilbert symbol defines a symmetric, nondegenerate bilinear form on $F^\times/F^{\times 2}$.*

Proof. Let $K := F[x]/(x^2 - a)$. The Hilbert symbol gives a well-defined map of sets

$$\begin{aligned} F^\times/F^{\times 2} &\to \{\pm 1\} \\ b &\mapsto (a, b)_F \end{aligned}$$

and we may conclude as in Lemma 12.4.6 if we show that this is a nontrivial group homomorphism.

First we show it is nontrivial. By Corollary 13.4.5, the field K embeds in the division quaternion algebra B, so by Exercise 2.5, there exists b such that $B \simeq (a, b \mid F)$, whence $(a, b)_F = -1$.

Next, we show it is a homomorphism. We appeal to Main Theorem 5.4.4. We have $(a, b)_F = 1$ if and only if $b \in \mathrm{Nm}_{K|F}(K^\times)$. So we reduce to showing that if $(a, b)_F = (a, b')_F = -1$ for $b, b' \in F^\times$, then $(a, bb')_F = 1$. But by Corollary 7.7.6, since there is a *unique* division quaternion algebra, we conclude that $b/b' \in \mathrm{Nm}_{K|F}(K^\times)$; thus $bb' = (b')^2(b/b') \in \mathrm{Nm}_{K|F}(K^\times)$ and $(a, bb' \mid F) \simeq \mathrm{M}_2(F)$ so $(a, bb')_F = 1$, as claimed. □

Remark 13.4.7. The proof of Corollary 13.4.6 (pairing with any $F^\times/F^{\times 2}$) shows that

$$F^\times/\,\mathrm{Nm}_{K|F}(K^\times) \simeq \mathbb{Z}/2\mathbb{Z}. \tag{13.4.8}$$

Conversely, if we know (13.4.8) then the properties of the Hilbert symbol are immediate. Although it was not hard to prove (13.4.8) when $\operatorname{char} k \neq 2$, to establish its truth when $\operatorname{char} k = 2$, one is led to study higher ramification groups (e.g. Serre [Ser79, Chapter XV]) eventually leading to local class field theory.

The norm groups played an important role in the proof above, so we conclude by recording the image of the reduced norm $\operatorname{nrd}(B_v^\times) \subseteq F_v^\times$.

Lemma 13.4.9. *We have*

$$\operatorname{nrd}(B^\times) = \begin{cases} \mathbb{R}_{>0}^\times, & \text{if } B \simeq \mathbb{H}; \\ F^\times, & \text{otherwise.} \end{cases}$$

Moreover, if F is nonarchimedean and $O \subseteq B$ is a maximal R-order, then $\operatorname{nrd}(O^\times) = R^\times$.

Proof. If $B \simeq \mathrm{M}_2(F)$ is split, then $\operatorname{nrd}(B^\times) = \det(\mathrm{GL}_2(F)) = F^\times$. So suppose B is a division algebra. If $B \simeq \mathbb{H}$ then $\operatorname{nrd}(B^\times) = \mathbb{R}_{>0}^\times$, so we suppose F is nonarchimedean. Then $B \simeq (K, \pi \mid F)$ where as above K is the unramified quadratic extension of F and π is a uniformizer. But $F^\times = R^\times \times \langle \pi \rangle$, and $\operatorname{nrd}(K^\times) = \operatorname{Nm}_{K|F}(K^\times) = R^\times \pi^{2\mathbb{Z}}$ and $\operatorname{nrd}(j) = \pi$. The result then follows by multiplicativity of the norm.

The second statement follows similarly: if $B \simeq \mathrm{M}_2(F)$ then $O \simeq \mathrm{M}_2(R)$ and $\operatorname{nrd}(O^\times) = \det(\mathrm{GL}_2(R)) = R^\times$; otherwise $O \simeq (S, \pi \mid R)$ where S is the ring of integers of K, and $\operatorname{nrd}(S^\times) = \operatorname{Nm}_{K|F}(S^\times) \supseteq R^\times$ and again $\operatorname{nrd}(j) = \pi$. □

13.5 Some topology

In this section, we dive into the basic topological adjectives relevant to the objects we have seen and that will continue to play an important role. Throughout, let F be a local field.

13.5.1. F is a locally compact topological field (by definition) but F is not itself compact. The subgroup $F^\times = F \smallsetminus \{0\}$ is given the topology induced from the embedding

$$\begin{aligned} F^\times &\hookrightarrow F \times F \\ x &\mapsto (x, x^{-1}); \end{aligned}$$

it turns out here that this coincides with the subspace topology $F^\times \subseteq F$ (see Exercise 13.14(a)). Visibly, $F^\times$ is open in F so $F^\times$ is locally compact.

If F is nonarchimedean, with valuation ring R and valuation v, then $F^\times$ is totally disconnected and further

$$R^\times = \{x \in R : v(x) = 0\} \subset R$$

is closed so is a topological abelian group that is compact (and totally disconnected).

Now let B be a finite-dimensional F-algebra.

13.5.2. As an F-vector space, B has a unique topology compatible with the topology on F as all norms on a topological vector F-space extending the norm on F are equivalent (the sup norm is equivalent to the sum of squares norm, etc.): see Exercise 13.12. In particular, two elements are close in the topology on B if and only if their coefficients are close with respect to a (fixed) basis: for example, two matrices in $\mathrm{M}_n(F)$ are close if and only if all of their coordinate entries are close. (Of course, the precise notion of "close" depends on the choice of norm.) Consequently, B is a complete, locally compact topological ring, taking a compact neighborhood in each coordinate.

13.5.3. The group $B^\times$ is a topological group, with the topology given by the embedding $B^\times \ni \alpha \mapsto (\alpha, \alpha^{-1}) \in B \times B$. This topology coincides with the subspace topology (see Exercise 13.14(b)). From this, we can see that $B^\times$ is locally compact: the norm $\mathrm{Nm}_{B|F} \colon B^\times \to F^\times$ is a continuous map, so $B^\times = \mathrm{Nm}_{B|F}^{-1}(F^\times)$ is open in B, and an open subset of a Hausdorff, locally compact space is locally compact in the subspace topology (Exercise 13.14(c)).

Example 13.5.4. If $B = \mathrm{M}_n(F)$, then $B^\times = \mathrm{GL}_n(F)$ is locally compact: a closed, bounded neighborhood that avoids the locus of matrices with determinant 0 is a compact neighborhood. When F is archimedean, this is quite visual: a matrix of nonzero determinant is at some finite distance away from the determinant zero locus! Note however that $\mathrm{GL}_n(F)$ is not itself compact: already $F^\times = \mathrm{GL}_1(F)$ is not compact.

Now suppose F is nonarchimedean with valuation v and valuation ring R.

13.5.5. We claim that R is the maximal compact subring of F. Indeed, $x \in F$ lies in a compact subring if and only if $v(x) \geq 0$ if and only if x is integral over R. The only new implication here is the statement that if $v(x) < 0$ then x does not lie in a compact subring, and that is because the sequence $x_n = x^n$ does not have a convergent subsequence as $|x_n| \to \infty$.

Next, let O be an R-order in B.

13.5.6. Choosing an R-basis, we have an isomorphism $O \simeq R^n$, and this isomorphism is also a homeomorphism. Therefore, O is compact as the Cartesian power of a compact set. The group $O^\times$ is therefore also compact because it is closed: for $\gamma \in O$, we have $\gamma \in O^\times$ if and only if $\mathrm{Nm}_{B|F}(\gamma) \in R^\times$, the norm map is continuous, and $R^\times = \{x \in R : v(x) = 0\} \subseteq R$ is closed.

Example 13.5.7. For $R = \mathbb{Z}_p \subseteq F = \mathbb{Q}_p$ and $B = \mathrm{M}_n(\mathbb{Q}_p)$, the order $O = \mathrm{M}_n(\mathbb{Z}_p)$ is compact (neighborhoods of a matrix can be taken as neighbhoods in each coordinate) and the subgroup $O^\times = \mathrm{GL}_n(\mathbb{Z}_p)$ is compact: there is no way to "run off to infinity", either in a single coordinate or via the determinant.

13.5.8. Suppose $B = D$ is a division algebra. Then the valuation ring O is the maximal compact subring of B, for the same reason as in the commutative case (see 13.5.5, details requested in Exercise 13.17(a)). There is a filtration

$$O \supset P \supset P^2 \supset \dots$$

giving rise to a filtration

$$O^\times \supset 1 + P \supset 1 + P^2 \supset \dots. \tag{13.5.9}$$

As in the second proof of Main Theorem 12.3.2, the quotient O/P is a finite extension of the finite residue field k, so $(O/P)^\times$ is a finite cyclic group. The maximal two-sided ideal P is principal, generated by an element j of minimal valuation, and multiplication by j^n gives an isomorphism $O/P \xrightarrow{\sim} P^n/P^{n+1}$ of k-vector spaces (or abelian groups) for all $n \geq 1$.

Furthermore, for each $n \geq 1$, there is an isomorphism of groups

$$\begin{aligned} O/P \simeq P^n/P^{n+1} &\xrightarrow{\sim} (1+P^n)/(1+P^{n+1}) \\ \alpha &\mapsto 1+\alpha. \end{aligned} \tag{13.5.10}$$

Therefore, $O^\times = \varprojlim_n (O/P^n)^\times$ is a projective limit of solvable groups, also called a **prosolvable** group.

Example 13.5.11. If B is a division quaternion algebra over $\mathbb{Q}_p$, with valuation ring O and maximal ideal P, then the filtration (13.5.9) has quotients isomorphic to $O/P \simeq \mathbb{F}_{p^2}$.

13.5.12. We will also want to consider norm 1 groups; for this, we suppose that B is a semisimple algebra. Let

$$B^1 := \{\alpha \in B : \operatorname{nrd}(\alpha) = 1\};$$

some authors also write $\mathrm{SL}_1(B) := B^1$. Then B^1 is a closed subgroup of $B^\times$, since the reduced norm is continuous.

If B is a division ring and F is archimedean, then $B \simeq \mathbb{H}$ and $B^1 \simeq \mathbb{H}^1 \simeq \mathrm{SU}(2)$ is compact (it is identified with the 3-sphere in $\mathbb{R}^4$). In a similar way, if B is a division ring and F is nonarchimedean, then B^1 is compact: for B has a valuation v and valuation ring O, and if $\alpha \in B$ has $\operatorname{nrd}(\alpha) = 1$ then $v(\alpha) = 0$ and $\alpha \in O$, and consequently $B^1 \subseteq O^\times$ is closed in a compact set so compact.

If B is not a division ring, then either B is the product of two algebras or B is a matrix ring over a division ring, and in either case B^1 is not compact.

Remark 13.5.13. The locally compact division algebras over a nonarchimedean field are necessarily totally disconnected. On the other hand, it is a theorem of Pontryagin [War89, Theorem 27.2] that if A is a *connected* locally compact division ring, then A is isomorphic as a topological ring to either $\mathbb{R}$, $\mathbb{C}$, or $\mathbb{H}$.

Exercises

1. Let $B := \left(\dfrac{-1,-1}{\mathbb{Q}_2}\right)$.

   (a) Show that B is a division ring that is complete with respect to the discrete valuation w defined by $w(t + xi + yj + zij) = v(t^2 + x^2 + y^2 + z^2)$ for $t, x, y, z \in \mathbb{Q}_2$.

(b) Prove that

$$O := \mathbb{Z}_2 \oplus \mathbb{Z}_2 i \oplus \mathbb{Z}_2 j \oplus \mathbb{Z}_2 \frac{1+i+j+ij}{2} \subset B$$

is the valuation ring of B.

2. Let B be a division quaternion algebra over a nonarchimedean local field F. Give another proof that the unramified quadratic extension K of F embeds in B as follows.

   Suppose it does not: then for all $\alpha \in O$, the extension $F(\alpha) \supseteq F$ is ramified, so there exists $a \in R$ such that $\alpha - a \in P \cap K(\alpha)$; let $P = jO$ and write $\alpha = \alpha_0 = a + j\alpha_1$, and iterate to conclude that $\alpha = \sum_{n=0}^{\infty} a_n j^n$ with $a_n \in R$. But $F(j)$ is complete so $O \subseteq F(j)$, a contradiction.

3. Let F be a local field with $F \not\simeq \mathbb{C}$, let K the unramified (separable) quadratic extension of F (take $K = \mathbb{C}$ if $F \simeq \mathbb{R}$), and let $\langle\sigma\rangle = \operatorname{Gal}(K \mid F)$, so that σ is the standard involution on K. Let B be a division quaternion algebra B over F.

   Show that
   $$B \simeq \left\{ \begin{pmatrix} a & b \\ \pi\sigma(b) & \sigma(a) \end{pmatrix} : a, b \in K \right\} \subseteq \mathrm{M}_2(K).$$

   [Hint: Compute the regular representation 2.3.8.] Identify the maximal order O its maximal ideal J under this identification.

4. Let B be a division quaternion algebra over F. Show that $\alpha \in B$ is integral over R if and only if $\operatorname{nrd}(\alpha), \operatorname{nrd}(\alpha+1) \in R$ if and only if $w(\alpha), w(\alpha+1) \geq 0$, where w is the valuation on B.

5. Extend Theorem 13.3.11 as follows. Let R be a complete DVR with field of fractions F, and let B be a quaternion division algebra over F. Show that $B \simeq \left(\dfrac{K, b}{F}\right)$ where $K \supseteq F$ is an unramified separable quadratic extension of F and $b \notin \operatorname{Nm}_{K|F}(K^\times)$.

6. Let F be a nonarchimedean local field with residue field k having $\operatorname{char} k \neq 2$, and let $K \supseteq F$ be a separable quadratic field extension.

   a) Let $b \in F^\times$. Show that if K is unramified then $b \in \operatorname{Nm}_{K|F}(K^\times)$ if and only if $v(b)$ is even; and if $K = F(\sqrt{a})$ is ramified, then $b \in \operatorname{Nm}_{K|F}(K^\times)$ if and only if $b = c^2$ or $b = -ac^2$ for some $c \in F^{\times 2}$.

   b) Deduce $[F^\times : \operatorname{Nm}_{K|F}(K^\times)] = 2$ and Corollary 13.4.6 in this case.

7. Let $R := \mathbb{Q}[[t]]$ be the ring of formal power series over $\mathbb{Q}$; then R is a complete DVR with fraction field $F = \mathbb{Q}((t))$, the Laurent series over $\mathbb{Q}$. Let $B_0 := (a, b \mid \mathbb{Q})$ be a division quaternion algebra over $\mathbb{Q}$, and let $B := B_0 \otimes_{\mathbb{Q}} F = (a, b \mid F)$. Show that B is a division quaternion algebra over F, with valuation ring $O := R + Ri + Rj + Rij$.

8. Let B be a division quaternion algebra over a nonarchimedean local field F, and let O be the valuation ring.

   (a) Show that every one-sided (left or right) ideal of O is a power of the maximal ideal P and hence is two-sided.

   (b) Let
   $$[O, O] := \langle \alpha\beta - \beta\alpha : \alpha, \beta \in O \rangle$$
   be the commutator ideal $[O, O]$ of O, the two-sided ideal generated by commutators of elements of O. Show that $P = [O, O]$.

9. Let F be a nonarchimedean local field, let $B = \mathrm{M}_2(F)$ and $O = \mathrm{M}_2(R)$. Show that there are $q + 1$ right O-ideals of norm $\mathfrak{p}$ corresponding to the elements of $\mathbb{P}^1(k)$ or equivalently the lines in k^2.

10. Give another proof of Lemma 13.4.9 using quadratic forms.

11. Let F be a nonarchimedean local field with valuation ring R, maximal ideal $\mathfrak{p}$, and residue field k. Let D be a division algebra over F with $\dim_F D = n^2$, with valuation ring O and maximal two-sided ideal P. Show that O/P is finite extension of k of degree n, and $P^n = \mathfrak{p}O$ (cf. Remark 13.4.3).

12. Show that (13.5.10) is an isomorphism of (abelian) groups.

13. Let F be a field with absolute value $|\ |$ and V a finite-dimensional F-vector space.

    (a) Let $x_1, \dots, x_n$ be a basis for V, and define
    $$\|a_1 x_1 + \dots + a_n x_n\| := \max(|a_1|, \dots, |a_n|)$$
    for $a_i \in F$. Show that V is a metric space with distance $d(x, y) = \|x - y\|$.

    (b) Show that the topology on V is independent of the choice of basis in (a).

    (c) Finally, show that if F is complete with respect to $|\ |$, then V is also complete.

14. Let F be a topological field. Show that the coarsest topology (fewest open sets) in which multiplication on $\mathrm{M}_n(F)$ is continuous is given by the coordinate topology.

15. Let F be a local field.

    (a) The group $F^\times$ has the structure of topological group under the embedding $x \mapsto (x, x^{-1}) \in F \times F$ (under the subspace topology in $F \times F$). Show that this topology coincides with the subspace topology $F^\times \subseteq F$.

    (b) More generally, let B be a finite-dimensional F-algebra. Show that the the topology on $B^\times$ induced by $\alpha \mapsto (\alpha, \alpha^{-1}) \in B \times B$ coincides with the subspace topology $B^\times \subseteq B$.

    (c) Show that an open subset of a Hausdorff, locally compact space is locally compact in the subspace topology.

16. Let F be a finite extension of $\mathbb{Q}_2$. Show that $(-1, -1)_F = (-1)^{[F:\mathbb{Q}_2]}$.

17. Finish the proof of Lemma 13.2.9.

18. Let D be a division algebra over a nonarchimedean local field F. We recall (see 13.5.2) that D is a complete, locally compact topological ring.

    (a) Verify (as in 13.5.5) that O is the maximal compact subring of B.

    (b) Show that $B^\times/F^\times$ is a compact topological group.

This chapter is licensed under the terms of the Creative Commons Attribution-NonCommercial 4.0 International License (http://creativecommons.org/licenses/by-nc/4.0/), which permits any noncommercial use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license and indicate if changes were made.

The images or other third party material in this chapter are included in the chapter's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the chapter's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.

Chapter 14

Quaternion algebras over global fields

In this chapter, we discuss quaternion algebras over global fields and characterize them up to isomorphism.

14.1 ▹ Ramification

To motivate the classification of quaternion algebras over $\mathbb{Q}$, we consider by analogy a classification of quadratic fields. We restrict to the following class of quadratic fields for the best analogy.

Definition 14.1.1. A quadratic field $F = \mathbb{Q}(\sqrt{d})$ of discriminant $d \in \mathbb{Z}$ is **mildly ramified** if $8 \nmid d$.

A quadratic field F is mildly ramified if and only if $F = \mathbb{Q}(\sqrt{m})$ where $m \neq 1$ is odd and squarefree; then $d = m$ or $d = 4m$ according as $m = 1, 3 \pmod 4$.

Let $F = \mathbb{Q}(\sqrt{d})$ be a mildly ramified quadratic field of discriminant $d \in \mathbb{Z}$ and let R be its ring of integers. A prime p ramifies in F, i.e. $pR = \mathfrak{p}^2$ for a prime ideal $\mathfrak{p} \subset R$, if and only if $p \mid d$.

But a discriminant d can be either positive or negative; to put this bit of data on the same footing, we define the set of **places** of $\mathbb{Q}$ to be the primes together with the symbol ∞, and we make the convention that ∞ ramifies in F if $d < 0$ and is unramified if $d > 0$. Let $F = \mathbb{Q}(\sqrt{d})$ be a mildly ramified quadratic field, and let $\operatorname{Ram}(F)$ be the set of places that ramify in F. The set $\operatorname{Ram}(F)$ determines F up to isomorphism, since the discriminant of F is the product of the odd primes in $\operatorname{Ram}(F)$, multiplied by 4 if $2 \in \operatorname{Ram}(F)$ and by -1 if $\infty \in \operatorname{Ram}(F)$. (For bookkeeping reasons, in this context it would probably therefore be better to consider 4 and -1 as primes, but we will resist the inducement here.) However, not every finite set of places Σ occurs: the product d corresponding to Σ is a discriminant if and only if $d \equiv 0, 1 \pmod 4$. We call this a **parity condition** on the set of ramifying places of a mildly ramified quadratic field:

$$2 \in \Sigma \quad \leftrightarrow \quad \text{there are an } \textit{odd} \text{ number of places in } \Sigma \text{ congruent to } -1 \pmod 4$$

with the convention that ∞ is congruent to $-1 \pmod 4$.

© The Author(s) 2021

J. Voight, *Quaternion Algebras*, Graduate Texts in Mathematics 288,
https://doi.org/10.1007/978-3-030-56694-4_14

Note that if Σ is a finite subset of places of $\mathbb{Q}$ and $2 \notin \Sigma$, then precisely one of either Σ or $\Sigma \cup \{\infty\}$ satisfies the parity condition; accordingly, if we define $m(\Sigma)$ to be the product of all odd primes in Σ multiplied by -1 if $\infty \in \Sigma$, then we can recover Σ from $m(\Sigma)$.

We have proven the following result.

Lemma 14.1.2. *The maps $F \mapsto \operatorname{Ram}(F)$ and $\Sigma \mapsto m(\Sigma)$ furnishes a bijection*

$$\left\{\begin{matrix}\text{Mildly ramified quadratic fields}\\ \mathbb{Q}(\sqrt{d}) \text{ up to isomorphism}\end{matrix}\right\} \leftrightarrow \left\{\begin{matrix}\text{Finite subsets of places of } \mathbb{Q}\\ \text{satisfying the parity condition}\end{matrix}\right\}$$
$$\leftrightarrow \left\{\begin{matrix}\text{Squarefree odd integers}\\ m \neq 1\end{matrix}\right\}.$$

This classification procedure using sets of ramifying primes and discriminants works as well for quaternion algebras over $\mathbb{Q}$. Let B be a quaternion algebra over $\mathbb{Q}$. When is a prime p ramified in B? In Chapter 12, we saw that the completion $B_p = B \otimes_{\mathbb{Q}} \mathbb{Q}_p$ is either a division ring or the matrix ring $\mathrm{M}_2(\mathbb{Q}_p)$. Further, when B_p is a division ring, the valuation ring $O_p \subset B_p$ is the unique maximal order, and the unique maximal ideal $P_p \subset O_p$ satisfies $pO_p = P_p^2$. By analogy with the quadratic case, we say that a place v is **ramified** in B if the completion B_v is a division ring, and otherwise v is **unramified** (or **split**).

Let $B = \left(\dfrac{a,b}{\mathbb{Q}}\right)$. Without loss of generality, we may suppose $a, b \in \mathbb{Z}$. There are only finitely many places where B is ramified: by the calculation of the Hilbert symbol (12.4.12), if p is prime and $p \nmid 2ab$, then $(a,b)_{\mathbb{Q}_p} = 1$ and p is split in B. Therefore $\#\operatorname{Ram} B < \infty$.

We say that B is **definite** if $\infty \in \operatorname{Ram} B$ and B is **indefinite** otherwise. By definition, B is definite if and only if $B_\infty := B \otimes_{\mathbb{Q}} \mathbb{R} = \left(\dfrac{a,b}{\mathbb{R}}\right) \simeq \mathbb{H}$ if and only if $a, b < 0$ (Exercise 2.4).

Let $\operatorname{Ram} B$ be the set of ramified places of B. Not every finite subset Σ of places can occur as $\operatorname{Ram} B$ for a quaternion algebra B. It turns out that the **parity condition** here is that we must have $\#\Sigma$ even. So again, if Σ is a finite set of primes, then precisely one of either Σ or $\Sigma \cup \{\infty\}$ can occur as $\operatorname{Ram} B$. We define the **discriminant** of B to be the product $\operatorname{disc} B$ of primes that ramify in B, so $\operatorname{disc} B$ is a squarefree positive integer. This notion of discriminant is admittedly strange; we relate it to perhaps more familiar notions in Chapter 15.

The main result of this chapter, specialized to the case $F = \mathbb{Q}$, is the following.

Main Theorem 14.1.3. *The maps $B \mapsto \operatorname{Ram} B$ and $\Sigma \mapsto \prod_{p \in \Sigma} p$ furnish bijections*

$$\left\{\begin{matrix}\text{Quaternion algebras over } \mathbb{Q}\\ \text{up to isomorphism}\end{matrix}\right\} \leftrightarrow \left\{\begin{matrix}\text{Finite subsets of places of } \mathbb{Q}\\ \text{of even cardinality}\end{matrix}\right\}$$
$$\leftrightarrow \{\, D \in \mathbb{Z}_{>0} \text{ squarefree}\,\}.$$

The composition of these maps is $B \mapsto \prod_{p \in \operatorname{Ram} B} p = \operatorname{disc} B$.

As previewed at the end of section 12.1, Main Theorem 14.1.3 is a *local-global principle* and provides a convenient way to test when quaternion algebras over $\mathbb{Q}$ are isomorphic: instead of working hard over $\mathbb{Q}$, we can just test for isomorphism over the local fields $\mathbb{Q}_p$ and $\mathbb{R}$.

We will spend the next two sections giving a self-contained proof of Main Theorem 14.1.3 following Serre [Ser73, Chapters III–IV], assuming two statements from basic number theory (quadratic reciprocity and the existence of primes in arithmetic progression), finishing the proof in section 14.3. Although the proofs presented do not seem to generalize beyond $F = \mathbb{Q}$, the argument is simple enough and its structure is good motivation for the more involved treatment in the Chapter ahead. (It is also comforting to see a complete proof in the simplest case.)

14.2 ▹ Hilbert reciprocity over the rationals

To begin, we look into the parity condition: it has a simple reformulation in terms of the Hilbert symbol (section 12.4). For a place v of $\mathbb{Q}$, let $\mathbb{Q}_v$ denote the completion of $\mathbb{Q}$ at the absolute value associated to v: if $v = p$ is prime, then $\mathbb{Q}_v = \mathbb{Q}_p$ is the field of p-adic numbers; if $v = \infty$, then $\mathbb{Q}_v = \mathbb{R}$. For $a, b \in \mathbb{Q}^\times$, we abbreviate $(a, b)_{\mathbb{Q}_v} = (a, b)_v$.

Proposition 14.2.1. (Hilbert reciprocity). *For all $a, b \in \mathbb{Q}^\times$, we have*

$$\prod_v (a, b)_v = 1, \tag{14.2.2}$$

where the product taken over all places v of $\mathbb{Q}$.

When p is odd and divides neither numerator nor denominator of a or b, we have $(a, b)_p = 1$, so the product (14.2.2) is well-defined. The following corollary is an equivalent statement.

Corollary 14.2.3. *Let B be a quaternion algebra over $\mathbb{Q}$. Then the set* Ram B *is finite of even cardinality.*

The law of Hilbert reciprocity, as it turns out, is a core premise in number theory: it is *equivalent* to the law of **quadratic reciprocity**

$$\left(\frac{p}{q}\right)\left(\frac{q}{p}\right) = (-1)^{\frac{p-1}{2}\frac{q-1}{2}} \tag{14.2.4}$$

for odd primes p, q together with the **supplement**

$$\left(\frac{-1}{p}\right) = (-1)^{\frac{p-1}{2}} \quad \text{and} \quad \left(\frac{2}{p}\right) = (-1)^{\frac{p^2-1}{8}} \tag{14.2.5}$$

for odd primes p.

We now give a proof of Hilbert reciprocity (Proposition 14.2.1), assuming the law of quadratic reciprocity and its supplement.

Proof of Proposition 14.2.1. Since each local Hilbert symbol is bilinear, it suffices to prove the statement when $a, b \in \mathbb{Z}$ are equal to either -1 or a prime number. The Hilbert symbol is also symmetric, so we may interchange a, b.

If $a = b = -1$, then $B = \left(\dfrac{a, b}{\mathbb{Q}}\right) = \left(\dfrac{-1, -1}{\mathbb{Q}}\right)$ is the rational Hamiltonians, and $(-1, -1)_\infty = (-1, -1)_2 = -1$ and $(-1, -1)_v = 1$ if $v \neq 2, \infty$, by the computation of the even Hilbert symbol (12.4.13). Similarly, the cases with $a = -1, 2$ follow from the supplement (14.2.5), and are requested in Exercise 14.1.

So we may suppose $a = p$ and $b = q$ are primes. If $p = q$ then $\left(\dfrac{p, p}{\mathbb{Q}}\right) \simeq \left(\dfrac{-1, p}{\mathbb{Q}}\right)$ and we reduce to the previous case, so we may suppose $p \neq q$. Since $p, q > 0$, we have $(p, q)_\infty = 1$. We have $(p, q)_\ell = 1$ for all primes $\ell \nmid 2pq$, and

$$(p, q)_p = (q, p)_p = \left(\frac{q}{p}\right) \quad \text{and} \quad (p, q)_q = \left(\frac{p}{q}\right)$$

by 12.4.12. Finally,

$$(p, q)_2 = -1 \text{ if and only if } p, q \equiv 3 \pmod 4$$

i.e., $(p, q)_2 = (-1)^{(p-1)(q-1)/4}$, again by the computation of the even Hilbert symbol (12.4.13). Thus the product becomes

$$\prod_v (p, q)_v = (-1)^{(p-1)(q-1)/4} \left(\frac{p}{q}\right)\left(\frac{q}{p}\right) = 1$$

by quadratic reciprocity. □

Hilbert reciprocity has several aesthetic advantages over the law of quadratic reciprocity. For one, it is simpler to write down! Also, Hilbert believed that his reciprocity law is a kind of analogue of Cauchy's integral theorem, expressing an integral as a sum of residues (Remark 14.6.4). The fact that a normalized product over all places is trivial also arises quite naturally: if we define for $x \in \mathbb{Q}^\times$ and a prime p the normalized absolute value

$$|x|_p := p^{-v_p(x)},$$

and $|x|_\infty$ the usual archimedean absolute value, then

$$\prod_v |x|_v = 1$$

by unique factorization in $\mathbb{Z}$; this is called the **product formula** for $\mathbb{Q}$, for obvious reasons.

From the tight relationship between quaternion algebras and ternary quadratic forms, we obtain the following corollary.

Corollary 14.2.6. *Let Q be a nondegenerate ternary quadratic form over $\mathbb{Q}$. Then the set of places v such that Q_v is anisotropic is finite and of even cardinality.*

In particular, by Corollary 14.2.6, if Q_v is isotropic for all but one place v of $\mathbb{Q}$, then Q_v is in fact isotropic for all places v.

Proof. In the bijection implied by Main Theorem 5.2.5, the quadratic form Q corresponds to a quaternion algebra $B = (a, b \mid \mathbb{Q})$, and by Main Theorem 5.4.4, Q is isotropic if and only if B is split if and only if $(a, b)_{\mathbb{Q}} = 1$. By functoriality, the same is true over each completion $\mathbb{Q}_v$ for v a place of $\mathbb{Q}$, and therefore the set of places v where Q_v is isotropic is precisely the set of ramified places in B. The result then follows by Hilbert reciprocity. □

To conclude this section, we show that every allowable product of Hilbert symbols is obtained.

Proposition 14.2.7. *Let Σ be a finite set of places of $\mathbb{Q}$ of even cardinality. Then there exists a quaternion algebra B over $\mathbb{Q}$ with* $\operatorname{Ram} B = \Sigma$.

Remark 14.2.8. Albert [Alb34, Theorem 2, Theorem 3] already sought to simplify the presentation of a quaternion algebra by a series of transformations, the content of which is contained in Proposition 14.2.7; this was further investigated by Latimer [Lat35].

Just as with Hilbert reciprocity, Proposition 14.2.7 touches on a deep statement in number theory concerning primes, due to Dirichlet.

Theorem 14.2.9. (Infinitude of primes in arithmetic progression). *Given $a, n \in \mathbb{Z}$ coprime, there are infinitely many primes $p \equiv a \pmod{n}$.*

Proof. See e.g. Serre [Ser73, Chapter VI] or Apostol [Apo76, Chapter 7]. We will prove this theorem in Exercise 26.11 as a consequence of the analytic class number formula. □

Remark 14.2.10. Theorem 14.2.9 seems to require analysis. (For algebraic proofs in special cases, see e.g., Neukirch [Neu99, Exercise I.10.1] and Lenstra–Stevenhagen [LS91].) Ram Murty [Mur88] showed that a “Euclidean proof” of the infinitude of primes $p \equiv a \pmod{n}$ is possible if and only if $a^2 \equiv 1 \pmod{n}$, and Paul Pollack [Pol2010] has shown that Schnizel’s Hypothesis H gives a heuristic for this. This crucial role played by analytic methods motivates part III of this monograph.

We now prove Proposition 14.2.7 assuming Theorem 14.2.9.

Proof. Let $D := \prod_{p \in \Sigma} p$ be the product of the primes in Σ, and let $u := -1$ if $\infty \in \Sigma$ and $u := 1$ otherwise. Let $D^\diamond := uD$. We consider quaternion algebras of the form

$$B = \left(\frac{q^\diamond, D^\diamond}{\mathbb{Q}}\right)$$

with $q^\diamond = uq$ (and q prime) chosen to satisfy certain congruence conditions ensuring that $\operatorname{Ram} B = \Sigma$. To this end, we seek a prime q such that

$$\left(\frac{q^\diamond}{p}\right) = -1 \text{ for all odd } p \mid D \tag{14.2.11}$$

and

$$q^\diamond \equiv \begin{cases} 1 \pmod 8, & \text{if } 2 \nmid D; \\ 5 \pmod 8, & \text{if } 2 \mid D. \end{cases} \tag{14.2.12}$$

There exists a prime satisfying the conditions (14.2.11)–(14.2.12) by Theorem 14.2.9, since the condition to be a quadratic nonresidue is a congruence condition on $q^\diamond$ and hence on q modulo p.

We now verify that B has $\operatorname{Ram} B = \Sigma$. We have $(q^\diamond, D^\diamond)_\infty = u$ by choice of signs and $(q^\diamond, D^\diamond)_p = 1$ for all $p \nmid 2dq$. We compute that

$$(q^\diamond, D^\diamond)_p = \left(\frac{q^\diamond}{p}\right) = -1 \quad \text{for all odd } p \mid D$$

by (14.2.11). For $p = 2$, we find that $(q^\diamond, D^\diamond)_2 = -1$ or $(q^\diamond, D^\diamond)_2 = 1$ according as $2 \mid D$ or not by the computation of the even Hilbert symbol (12.4.13). This shows that

$$\Sigma \subseteq \operatorname{Ram} B \subseteq \Sigma \cup \{q\}.$$

The final symbol $(q^\diamond, D^\diamond)_q$ is determined by Hilbert reciprocity (Proposition 14.2.1): since $\#\Sigma$ is already even, we must have $(q^\diamond, D^\diamond)_q = 1$. Therefore the quaternion algebra $B := \left(\dfrac{q^\diamond, D^\diamond}{\mathbb{Q}}\right)$ has $\Sigma = \operatorname{Ram} B$. □

Example 14.2.13. Let $B = (a, b \mid \mathbb{Q})$ be a quaternion algebra of prime discriminant $D = p$ over $\mathbb{Q}$. Then:

(i) For $D = p = 2$, we take $a = b = -1$;
(ii) For $D = p \equiv 3 \pmod 4$, we take $b = -p$ and $a = -1$;
(iii) For $D = p \equiv 1 \pmod 4$, we take $b = -p$ and $a = -q$ where $q \equiv 3 \pmod 4$ is prime and $\left(\dfrac{q}{p}\right) = -1$.

In case (iii), by qudaratic reciprocity $\left(\dfrac{-p}{q}\right) = -\left(\dfrac{q}{p}\right) = 1$ so indeed B is not ramified at p. In the proof of Theorem 14.2.7 above, we would have required the more restrictive condition $q \equiv 3 \pmod 8$, but we can look again at the table of even Hilbert symbols (12.4.16): since $b = -p = -1, 3 \pmod 8$, we may take $a = -q = 1, -3 \pmod 8$ freely, so $q \equiv 3 \pmod 4$.

Similarly, for discriminant D the product of two (distinct) primes:

(i) For $D = 2p$ with $p \equiv 3 \pmod 4$, we take $a = -1$ and $b = p$;
(ii) For $D = 2p$ with $p \equiv 5 \pmod 8$, we take $a = 2$ and $b = p$;
(iii) For $D = pq$ with $p \equiv q \equiv 3 \pmod 4$, we take $a = -1$ and $b = pq$;
(iv) For $D = pq$ with $p \equiv 1 \pmod 4$ or $q \equiv 1 \pmod 4$ and $\left(\dfrac{q}{p}\right) \neq 1$, we take $a = p$ and $b = q$.

For other explicit presentations of quaternion algebras over $\mathbb{Q}$ with specified discriminant, see Alsina–Bayer [AB2004, §1.1.2]. See Example 15.5.7 for some explicit maximal orders.

14.3 ▹ Hasse–Minkowski theorem over the rationals

To complete the proof of Main Theorem 14.1.3, we now show that the map $B \mapsto \operatorname{Ram} B$ is injective on isomorphism classes.

Proposition 14.3.1. *Let B, B' be quaternion algebras over $\mathbb{Q}$. Then the following are equivalent:*

(i) $B \simeq B'$;
(ii) $\operatorname{Ram} B = \operatorname{Ram} B'$;
(iii) $B_v \simeq B'_v$ *for all places* $v \in \operatorname{Pl}(\mathbb{Q})$; *and*
(iv) $B_v \simeq B'_v$ *for all but one place* v.

The statement of Proposition 14.3.1 is a *local-global principle*: the global isomorphism class is determined by the local isomorphism classes.

Corollary 14.3.2. *Let B be a quaternion algebra over $\mathbb{Q}$. Then $B \simeq \mathrm{M}_2(\mathbb{Q})$ if and only if $B_p \simeq \mathrm{M}_2(\mathbb{Q}_p)$ for all primes p.*

Proof. Apply Proposition 14.3.1 (i) ⇔ (iv) with $B' = \mathrm{M}_2(\mathbb{Q})$ checking all but the archimedean place. □

By the equivalence between quaternion algebras and quadratic forms (see Chapter 5, specifically section 5.2), the statement of Proposition 14.3.1 is equivalent to the statement that a ternary quadratic form over $\mathbb{Q}$ is isotropic if and only if it is isotropic over all (but one) completions. In fact, the more general statement is true—and again we come in contact with a deep result in number theory.

Theorem 14.3.3. (Hasse–Minkowski). *Let Q be a quadratic form over $\mathbb{Q}$. Then Q is isotropic if and only if Q_v is isotropic for all places v of $\mathbb{Q}$.*

We will prove the Hasse–Minkowski theorem by induction on the number of variables. Of particular interest is the case of (nondegenerate) ternary quadratic forms, for which we have the following theorem of Legendre.

Theorem 14.3.4. (Legendre). *Let $a, b, c \in \mathbb{Z}$ be nonzero, squarefree integers that are relatively prime in pairs. Then the quadratic form*

$$ax^2 + by^2 + cz^2 = 0$$

has a nontrivial solution $x, y, z \in \mathbb{Q}$ if and only if a, b, c do not all have the same sign and

$$-ab,\ -bc,\ -ac \text{ are quadratic residues modulo } |c|,\ |a|,\ |b|, \text{ respectively.}$$

Proof. First, the conditions for solvability are necessary. The condition on signs is necessary for a solution in $\mathbb{R}$. If $ax^2 + by^2 + cz^2 = 0$ with $x, y, z \in \mathbb{Q}$ not all zero, then scaling we may suppose $x, y, z \in \mathbb{Z}$ satisfy $\gcd(x, y, z) = 1$; if $p \mid c$ then $p \nmid y$ (else $p \mid x$ and $p \mid z$, contradiction), so $(x/y)^2 \equiv (-b/a) \pmod{|c|}$ and $-ba$ is a quadratic residue modulo $|c|$; the other conditions hold by symmetry.

So suppose the conditions hold. Multiplying through and rescaling by squares (Exercise 14.8), we may suppose a, b are squarefree (but not necessarily coprime) and $c = -1$, and we seek a nontrivial solution to $ax^2 + by^2 = z^2$. If $a \in \mathbb{Q}^{\times 2}$, then we are done. Otherwise, we need to solve

$$\frac{z^2 - ax^2}{y^2} = b = \operatorname{Nm}_{\mathbb{Q}(\sqrt{a})/\mathbb{Q}}\left(\frac{z + x\sqrt{a}}{y}\right)$$

for $x, y, z \in \mathbb{Q}$ and $y \neq 0$, i.e., we need to show that b is a norm from $F = \mathbb{Q}(\sqrt{a})$. By hypothesis, a, b are not both negative and

$$b \text{ is a square modulo } |a| \text{ and } a \text{ is a square modulo } |b|. \tag{14.3.5}$$

We may also suppose $|a| \leq |b|$.

We use complete induction on $m = |a| + |b|$. If $m = 2$, then we must consider the equation $\pm x^2 \pm y^2 = z^2$ with the case both negative signs excluded, each of which has solutions. Now suppose that $m > 2$ so $|b| \geq 2$, and let $p \mid b$ be prime divisor. By hypothesis, there exist integers t, b' such that $t^2 = a + bb'$; taking a small residue, we may suppose $|t| < |b|/2$. Thus

$$bb' = t^2 - a = \operatorname{Nm}_{F/\mathbb{Q}}(t + \sqrt{a})$$

so bb' is a norm from F. Thus b is a norm if and only if b' is a norm. But

$$|b'| = \left|\frac{t^2 - a}{b}\right| \leq \frac{|b|}{4} + 1 < |b|$$

because $|b| \geq 2$.

Now write $b' = b''u^2$ with $b'', u \in \mathbb{Z}$ and b'' squarefree. Then $|b''| \leq |b'| < |b|$ and b'' is a norm if and only if b' is a norm. With these manipulations, we propagate the hypothesis that $|a|$ is a square modulo $|b''|$ and $|b''|$ is a square modulo $|a|$. Therefore, the induction hypothesis applies to the equation $ax^2 + b''y^2 = z^2$, and the proof is complete. □

Corollary 14.3.6. *Let Q be a nondegenerate ternary quadratic form over $\mathbb{Q}$. Then Q is isotropic if and only if Q_v is isotropic for all places v of $\mathbb{Q}$ (but one).*

Proof. If Q is isotropic, then Q_v is isotropic for all v. For the converse, suppose that Q_v is isotropic for all places v of $\mathbb{Q}$. As in the proof of Legendre's Theorem 14.3.4, we may suppose $Q(x, y, z) = ax^2 + by^2 - z^2$. The fact that Q is isotropic over $\mathbb{R}$ implies that a, b are not both negative. Now let $p \mid a$ be odd. The condition that Q_p is isotropic is equivalent to $(a, b)_p = (b/p) = 1$; putting these together, we conclude that b is a quadratic residue modulo $|a|$. The same holds for a, b interchanged, so (14.3.5) holds and the result follows. □

We are now in a position to complete the proof of the Hasse–Minkowski theorem.

Proof of Theorem 14.3.3. We follow Serre [Ser73, Theorem 8, §IV.3.2]. We may suppose that Q is nondegenerate in $n \geq 1$ variables. If $n = 1$, the statement is vacuous. If $n = 2$, the after scaling we may suppose $Q(x, y) = x^2 - ay^2$ with $a \in \mathbb{Q}^\times$; since Q_p is isotropic for all primes p, we have $a \in \mathbb{Q}_p^{\times 2}$ so in particular $v_p(a)$ is even for all primes p; since Q is isotropic at ∞, we have $a > 0$; thus by unique factorization $a \in \mathbb{Q}^{\times 2}$, and the result follows. If $n = 3$, the statement is proven in Corollary 14.3.6.

Now suppose $n \geq 4$. Write $Q = \langle a, b\rangle \boxplus -Q'$ where $Q' = \langle c_1, \dots, c_{n-2}\rangle$ and $a, b, c_i \in \mathbb{Z}$. Let $d = 2ab(c_1 \cdots c_{n-2}) \neq 0$. For each prime $p \mid d$, since Q is isotropic, there exists $t_p \in \mathbb{Q}_p^\times$ represented by both $\langle a, b\rangle$ and Q' in $\mathbb{Q}_p$. (This requires a small argument, see Exercise 6.14.) Similarly, there exists $t_\infty \in \mathbb{R}^\times$ represented by these forms in $\mathbb{R}$.

By another application of the infinitude of primes in arithmetic progression (Exercise 14.10), there exists $t \in \mathbb{Q}^\times$ such that:

(i) $t \in t_p \mathbb{Q}_p^{\times 2}$ for all primes $p \mid d$,
(ii) t and t_∞ have the same sign, and
(iii) $p \nmid t$ for all primes $p \nmid d$ except possibly for one prime $q \nmid d$.

Now the quadratic form $\langle a, b, -t\rangle$ is isotropic for all $p \mid d$ and at ∞ by construction and at all primes $p \nmid d$ except $p = q$ since $p \nmid abt$. Therefore, by case $n = 3$ (using the "all but one" in Corollary 14.3.6), the form $\langle a, b, -t\rangle$ is isotropic.

On the other side, if $n = 4$, then the form $\langle t\rangle \boxplus Q'$ is isotropic by the same argument. If $n \geq 5$, then we apply the induction hypothesis to Q': the hypothesis holds, since Q' is isotropic at ∞ and all $p \mid d$ by construction, and for all $p \nmid d$ the completion Q'_p is a nondegenerate form in ≥ 3 variables over $\mathbb{Z}_p$ so is isotropic by the results of section 12.3, using Hensel's lemma to lift a solution modulo the odd prime p.

Putting these two pieces together, we find that Q is isotropic over $\mathbb{Q}$. □

We conclude with the following consequence.

Corollary 14.3.7. *Let Q, Q' be quadratic forms over $\mathbb{Q}$ in the same number of variables. Then $Q \simeq Q'$ if and only if $Q_v \simeq Q'_v$ for all places v.*

Proof. The implication ($\Rightarrow$) is immediate. We prove ($\Leftarrow$) by induction on the number of variables, the case of $n = 0$ variables being clear. By splitting the radical (4.3.9), we may suppose that Q, Q' are nondegenerate. Let $a \in \mathbb{Q}^\times$ be represented by Q. Since $Q_v \simeq Q'_v$ the quadratic form $\langle -a\rangle \boxplus Q'$ is isotropic at v for all v, so Q' represents a (Lemma 5.4.3). In both cases, we can write $Q \simeq \langle -a\rangle \boxplus Q_1$ and $Q' \simeq \langle -a\rangle \boxplus Q'_1$ for quadratic forms Q_1, Q'_1 in one fewer number of variables. Finally, by Witt cancellation (Theorem 4.2.22), from $Q_v \simeq Q'_v$ we have $(Q_1)_v \simeq (Q'_1)_v$ for all v, so by induction $Q_1 \simeq Q'_1$, and thus $Q \simeq Q'$. □

We now officially complete our proofs.

Proof of Proposition 14.3.1. The implications (i) $\Rightarrow$ (ii) and (iii) $\Rightarrow$ (iv) are immediate. For the implication (ii) $\Rightarrow$ (iii): either $v \in \operatorname{Ram} B$, in which case $B_v \simeq B'_v$ is the unique division algebra over $\mathbb{Q}_v$ (Theorem 12.1.5), or $v \notin \operatorname{Ram} B$, in which case $B_v \simeq \mathrm{M}_2(\mathbb{Q}_v) \simeq B'_v$ by definition. For the implication (iv) $\Rightarrow$ (i), recalling

Theorem 5.1.1, by Corollary 14.3.6 applied to the ternary quadratic form associated to B, we conclude that this form is isotropic, which by Proposition 5.1.2 implies that $B \simeq \mathrm{M}_2(\mathbb{Q})$. □

Proof of Main Theorem 14.1.3. The map $B \mapsto \operatorname{Ram} B$ has the desired codomain, by Hilbert reciprocity (Proposition 14.2.1); it is surjective by Proposition 14.2.7; and it is injective by Corollaries 14.3.6 and 14.3.7. The second bijection (with squarefree integers) is immediate. □

To summarize these past few sections, the classification of quaternion algebras over $\mathbb{Q}$ embodies some deep statements in number theory: quadratic reciprocity (and its reformulation in Hilbert reciprocity), the Hasse–Minkowski theorem (the local-global principle for quadratic forms), and the proofs use the theorem of the infinitude of primes in arithmetic progression! It is a small blessing that we can make these essentially elementary arguments over $\mathbb{Q}$. In the more general case, we must dig more deeply.

For fun, we conclude this section with a consequence in number theory: Legendre's three-square theorem (cf. Lagrange's four-square theorem, Theorem 11.4.3, and Remark 11.4.4).

Theorem 14.3.8. (Legendre–Gauss). *An integer $n \geq 0$ can be written as the sum of three squares $n = x^2 + y^2 + z^2$ if and only if n is not of the form $n = 4^a(8b + 7)$ with $a, b \in \mathbb{Z}$.*

Proof. Looking modulo 8, we see that the provided condition is necessary (Exercise 14.3(a)). Conversely, suppose $n > 0$ is not of the form $n = 4^a(8b + 7)$, or equivalently that $-n \notin \mathbb{Q}_2^{\times 2}$ (Exercise 14.4). We may suppose $a = 0, 1$.

Let $B = (-1, -1 \mid \mathbb{Q})$ be the rational Hamiltonians. We have $\operatorname{Ram} B = \{2, \infty\}$, which is to say the associated ternary quadratic form $x^2 + y^2 + z^2$ is isotropic over $\mathbb{Q}_p$ for all odd primes p. Consider the quadratic form $Q(x, y, z, w) = x^2 + y^2 + z^2 - nw^2$. Then Q is isotropic over $\mathbb{R}$ since $n > 0$, and isotropic over all $\mathbb{Q}_p$ with p odd taking $w = 0$. The form is also isotropic over $\mathbb{Q}_2$ (Exercise 14.3), lifting a solution modulo 8 via Hensel's lemma. By the Hasse–Minkowski theorem (Theorem 14.3.3), Q is isotropic over $\mathbb{Q}$, so there exist $x, y, z, w \in \mathbb{Q}$ not all zero such that $x^2 + y^2 + z^2 = nw^2$. We must have $w \neq 0$ by positivity, and dividing through we get $x, y, z \in \mathbb{Q}$ not all zero such that $x^2 + y^2 + z^2 = n$. Let $\alpha = xi + yj + zij \in B$. Then $\alpha^2 + n = 0$ and $\alpha \in B$ is integral.

Let $\mathcal{O}' \subset B$ be a maximal order containing α, and let $\mathcal{O}$ be the Hurwitz order. By Proposition 11.3.7, $\mathcal{O}'$ is conjugate to $\mathcal{O}$; after conjugating, we may suppose $\alpha \in \mathcal{O}$. But $\operatorname{trd}(\alpha) = 0$, so necessarily $\alpha \in \mathbb{Z}\langle i, j\rangle$ and $x, y, z \in \mathbb{Z}$ with $\operatorname{nrd}(\alpha) = x^2 + y^2 + z^2 = n$ as desired. □

See also Exercise 14.5 for a variant of the proof of the three-square theorem staying in the language of quaternions.

14.4 Global fields

In this chapter and in many that remain, we focus on a certain class of fields of arithmetic interest: a **global field** is either a finite extension of $\mathbb{Q}$ (a **number field**) or of $\mathbb{F}_p(t)$ (a **function field**) for a prime p. Global fields are strongly governed by their completions with respect to nontrivial absolute values, which are local fields. Throughout this text, we will return to this theme that global behavior is governed by local behavior.

For the rest of this chapter, let F be a global field. We quickly introduce in this section some basic notions from algebraic number theory: for further reference, see e.g. Neukirch [Neu99, Chapters I–II], Cassels [Cas2010, Chapter II], or Janusz [Jan96, Chapter II].

Remark 14.4.1. When F is a function field, we will often insist that F is equipped with an inclusion $F_0 \hookrightarrow F$ where $F_0 \simeq \mathbb{F}_p(t)$ has pure transcendence degree 1 over $\mathbb{F}_p$. (For the geometrically inclined, this corresponds to a morphism $X \to \mathbb{P}^1$ of the associated curves.) Often this inclusion will not play a role, but it will be important to treat certain aspects uniformly with the number field case where there is only one inclusion $\mathbb{Q} \hookrightarrow F$.

14.4.2. The set of **places** of F is the set $\operatorname{Pl} F$ of equivalence classes of embeddings $\iota_v \colon F \to F_v$ where F_v is a local field and $\iota_v(F)$ is dense in F_v; two embeddings $\iota_v \colon F \to F_v$ and $\iota_v' \colon F \to F_v'$ are said to be **equivalent** if there is an isomorphism of topological fields $\phi \colon F_v \to F_v'$ such that $\iota_v' = \phi \circ \iota_v$.

14.4.3. Every valuation $v \colon F \to \mathbb{R} \cup \{\infty\}$, up to scaling, defines a place $\iota_v \colon F \to F_v$ where v is the completion of F with respect to the absolute value induced by v; we call such a place **nonarchimedean**, and using this identification we will write v for both the place of F and the corresponding valuation. For a nonarchimedean place v corresponding to a local field F_v, we denote by R_v its valuation ring, $\mathfrak{p}_v$ its maximal ideal, and k_v its residue field. If F is a function field, then all places of F are nonarchimedean. If F is a number field, a place $F \hookrightarrow \mathbb{R}$ is called a **real place** and a place $F \hookrightarrow \mathbb{C}$ (equivalent to its complex conjugate) is called a **complex place**. A real or complex place is **archimedean**.

14.4.4. Let $K \supseteq F$ be a finite, separable extension of fields, and let $v \in \operatorname{Pl} F$. We say that a place w of K is **above** v if $w|_F = v$, and we write $w \mid v$. The set of places w above v are obtained as follows: since K is separable, we have an isomorphism

$$K \otimes_F F_v \simeq K_1 \times \cdots \times K_r \tag{14.4.5}$$

where each $K_i \supseteq F_v$ is a finite extension of local fields. Indeed, writing $K = F[x]/(f(x))$ with $f(x) \in F[x]$ the minimal polynomial of a primitive element, we have

$$K \otimes_F F_v \simeq F_v[x]/(f(x)) \simeq F_v[x]/(f_1(x)) \times \cdots \times F_v[x]/(f_r(x))$$

where $f(x) = f_1(x) \cdots f_r(x) \in F_v[x]$ is the factorization of $f(x)$ into irreducibles in $F_v[x]$, distinct because f is separable. Thus each K_i is a local field by the classification

in Theorem 12.2.15, and the composition

$$K \hookrightarrow K \otimes_F F_v \to K_i$$

defines a place w_i of K above v. Conversely, every place w above v is equivalent to w_i for some i [Jan96, Chapter II, Theorem 5.1; Cas78, §9; Neu99, Chapter II, Proposition (8.3)].

We say that a nonarchimedean place v **ramifies** in K if there exists a place $w \mid v$ such that $K_w \supseteq F_v$ is ramified (see 13.2.3). Only finitely many places of F ramify in K.

A global field F has a set of **preferred** embeddings $\iota_v : F \hookrightarrow F_v$ corresponding to each place $v \in \operatorname{Pl} F$—equivalently, a preferred choice of absolute values $|\,|_v$ for each place $v \in \operatorname{Pl} F$—such that the **product formula** holds: for all $x \in F^\times$,

$$\prod_{v \in \operatorname{Pl} F} |x|_v^{m_v} = 1. \tag{14.4.6}$$

where $m_v = 2$ if v is complex and $m_v = 1$ otherwise. Admittedly, the extra exponents 2 for the complex places are annoying (see Remark 12.2.3)! Often what is done is to define **normalized absolute values** $\|x\|_v := |x|_v^{m_v}$ for $v \in \operatorname{Pl} F$, so then (14.4.6) becomes

$$\prod_{v \in \operatorname{Pl} F} \|x\|_v = 1. \tag{14.4.7}$$

Preferred absolute values are defined as follows.

14.4.8. The set of places $\operatorname{Pl}(\mathbb{Q})$ of $\mathbb{Q}$ consists of the archimedean real place, induced by the embedding $\mathbb{Q} \hookrightarrow \mathbb{R}$ and the usual absolute value $|x|_\infty$, and the set of nonarchimedean places indexed by the primes p given by the embeddings $\mathbb{Q} \hookrightarrow \mathbb{Q}_p$, with the preferred absolute value

$$|x|_p = p^{-v_p(x)}.$$

The statement of the product formula for $x \in \mathbb{Q}$ is

$$|x|_\infty \prod_p p^{-v_p(x)} = 1; \tag{14.4.9}$$

rearranging, (14.4.9) is equivalent to $\prod_p p^{v_p(x)} = |x|$, and this follows from unique factorization in $\mathbb{Z}$.

14.4.10. The set of places of $\mathbb{F}_p(t)$ is indexed by monic irreducible polynomials $f(t) \in \mathbb{F}_p[t]$ with preferred absolute value

$$|x(t)|_f = p^{-(\deg f)\operatorname{ord}_f(x)}$$

and $1/t$, the place at infinity, with preferred absolute value

$$|x(t)|_{1/t} = p^{\deg x},$$

where if $x = f/g$ is the ratio of relatively prime polynomials $f, g \in \mathbb{F}_p[t]$, then $\deg x := \max(\deg f, \deg g)$.

Then the statement of the product formula for $x(t) \in \mathbb{F}_p(t)$ is

$$p^{\deg x} \prod_f p^{-(\deg f) \operatorname{ord}_f(x)} = 1; \tag{14.4.11}$$

rearranging as over $\mathbb{Q}$, but now also taking the logarithm in base p, (14.4.9) is equivalent to $\sum_f (\deg f) \operatorname{ord}_f(x) = \deg x$ which follows from unique factorization in $\mathbb{F}_p[t]$.

14.4.12. More generally, let $K \supseteq F$ be a finite, separable extension of global fields. Let v be a place of F with a preferred absolute value and let w be a place of K above v. Then the preferred absolute value for w is the unique one extending v, namely

$$|x|_w = |\operatorname{Nm}_{K_w|F_v}(x)|_v^{1/[K_w:F_v]}$$

for $x \in K$. These absolute values fit together, with

$$\prod_{w|v} |x|_w = |\operatorname{Nm}_{K|F}(x)|_v \tag{14.4.13}$$

for all $x \in K$, a consequence of (14.4.5) [Jan96, Chapter II, Theorem 5.2; Cas78, §11, Theorem, p. 59; Neu99, Chapter II, Corollary (8.4)].

In particular, if F satisfies the product formula (14.4.6) with respect to preferred absolute values, then so does K, since

$$\prod_w |x|_w^{m_w} = \prod_v \left(\prod_{w|v} |x|_w^{m_w} \right) = \prod_v |\operatorname{Nm}_{K|F}(x)|_v^{m_v} = 1. \tag{14.4.14}$$

Remark 14.4.15. The definitions for the preferred absolute values are pretty dry—sorry! But we will see later that they are natural from the perspective of *Haar measure*: see section 29.3 and ultimately (29.6.3).

We will also make use of the following notation in many places in the text. Let F be a global field.

Definition 14.4.16. A set $S \subseteq \operatorname{Pl} F$ is **eligible** if S is finite, nonempty, and contains all archimedean places of F.

Definition 14.4.17. Let S be an eligible set of places. The **ring of S-integers** in F is the set

$$R_{(S)} := \{x \in F : v(x) \geq 0 \text{ for all } v \notin S\}. \tag{14.4.18}$$

A **global ring** is a ring of S-integers in a global field for an associated eligible set S.

The expression (14.4.18) makes sense, since if $v \notin S$ then by hypothesis v is nonarchimedean. When no confusion can result, we will abbreviate $R = R_{(S)}$ for a global ring R.

Example 14.4.19. If F is a number field and S consists only of the archimedean places in F then $R_{(S)}$ is the **ring of integers** in F, the integral closure of $\mathbb{Z}$ in F, also denoted $R_{(S)} = \mathbb{Z}_F$. If F is a function field, corresponding to a curve X, then $R_{(S)}$ is the ring of all rational functions with no poles outside S. (So in all cases, it is helpful to think of the ring $R_{(S)}$ as consisting of those elements of F with "no poles outside S".)

14.5 Ramification and discriminant

Let $R = R_{(S)}$ be a global ring, with $S \subset \operatorname{Pl} F$ eligible. Let B be a quaternion algebra over F.

Definition 14.5.1. Let $v \in \operatorname{Pl} F$. We say that B is **ramified** at v if $B_v = B \otimes_F F_v$ is a division ring; otherwise we say that B is **split** (or **unramified**) at v.

Let $\operatorname{Ram} B$ denote the set of ramified places of B.

If $v \in \operatorname{Pl} F$ is a nonarchimedean place, corresponding to a prime $\mathfrak{p}$ of R, we will also say that B is **ramified** at $\mathfrak{p}$ when B is ramified at v.

Remark 14.5.2. We use the term *ramified* for the following reason: if $B_{\mathfrak{p}}$ is a division ring with valuation ring $O_{\mathfrak{p}}$, then $\mathfrak{p}O_{\mathfrak{p}} = P^2$ for a two-sided maximal ideal P: see Theorem 13.3.11. (Eichler [Eic55-56, §1, Theorem 4] called them *characteristic* primes.)

Lemma 14.5.3. *The set* $\operatorname{Ram} B$ *of ramified places of* B *is finite.*

Proof. Write $B = (K, b \mid F)$. Since F has only finitely many archimedean places, we may suppose v is nonarchimedean. The extension $K \supseteq F$ is ramified at only finitely many places, so we may suppose that $K \supseteq F$ is unramified at v. Finally, $v(b) = 0$ for all but finitely many v, so we may suppose $v(b) = 0$. But then under these hypotheses, $B_v = (K_v, b \mid F_v)$ is split, by Corollary 13.4.1. □

Motivated by the fact that the discriminant of a quadratic field extension is divisible by ramifying primes, we make the following definition.

Definition 14.5.4. The R-**discriminant** of B is the R-ideal

$$\operatorname{disc}_R(B) = \prod_{\substack{\mathfrak{p} \in \operatorname{Ram} B \\ \mathfrak{p} \notin S}} \mathfrak{p} \subseteq R$$

obtained as the product of all primes $\mathfrak{p}$ of $R = R_{(S)}$ ramified in B.

Remark 14.5.5. When F is a number field and S consists of archimedean places only, so that $R = \mathbb{Z}_F$ is the ring of integers of F, we abbreviate $\operatorname{disc}_R(B) = \operatorname{disc} B$. The discriminant $\operatorname{disc}_R(B)$ discards information about primes in S: only $\operatorname{Ram} B$ records information about B that is independent of S.

Remark 14.5.6. One could make the same definitions when R is more generally a Dedekind domain. However, unless the residue fields of R are finite, this is not as useful a notion: see Exercise 14.14. (In some sense, this is because the Brauer group of $F = \operatorname{Frac} R$ is not as simply described as when F is a global field, viz. Remark 14.6.10.)

As usual, the archimedean places play a special role for number fields, so we make the following definition.

Definition 14.5.7. Let F be a number field. We say that B is **totally definite** if all archimedean places of F are ramified in B; otherwise, we say B is **indefinite**.

14.5.8. If v is a complex place, then v is necessarily split, since the only quaternion algebra over $\mathbb{C}$ is $\mathrm{M}_2(\mathbb{C})$; therefore, if B is a totally definite quaternion algebra over a number field F, then F is totally real.

14.6 Quaternion algebras over global fields

We now generalize Main Theorem 14.1.3 to the global field F, deducing results characterizing isomorphism classes of quaternion algebras. The main result is as follows.

Main Theorem 14.6.1. *Let F be a global field. Then the map $B \mapsto \operatorname{Ram} B$ gives a bijection*

$$\left\{ \begin{array}{c} \text{Quaternion algebras over } F \\ \text{up to isomorphism} \end{array} \right\} \leftrightarrow \left\{ \begin{array}{c} \text{Finite subsets of noncomplex places} \\ \text{of } F \text{ of even cardinality} \end{array} \right\}.$$

In other words, if B is a quaternion algebra over a global field, then the set of places of F where B is ramified is finite and of even cardinality, this set uniquely determines B up to isomorphism, and every such set occurs.

Proof. We give a proof in section 26.8, which itself relies on an analytic result (Theorem 26.8.19) proven in Chapter 29.

Alternatively, this statement can also be viewed a direct consequence of a (hard-earned) fundamental exact sequence in class field theory: see Remark 14.6.10. □

Recall the definition of the Hilbert symbol (as in section 12.4), computed explicitly for v an odd nonarchimedean place (12.4.9): for a place v of F, we abbreviate $(a,b)_{F_v} = (a,b)_v$. We also recall Lemma 14.5.3 that $(a,b)_v = 1$ for all but finitely many places v.

Corollary 14.6.2. (Hilbert reciprocity). *Let F be a global field with $\operatorname{char} F \neq 2$ and let $a, b \in F^\times$. Then*

$$\prod_{v \in \operatorname{Pl} F} (a,b)_v = 1. \tag{14.6.3}$$

Proof. Immediate from Main Theorem 14.6.1: Hilbert reciprocity is equivalent to the statement that $\#\operatorname{Ram} B$ is even. □

Remark 14.6.4. Stating the reciprocity law in the form (14.6.3) is natural from the point of view of the product formula (14.4.6). And Hilbert reciprocity can be rightly seen as a law of quadratic reciprocity for number fields (as we saw in section 14.2 for $F = \mathbb{Q}$). (For more, see Exercise 14.16.)

Hilbert saw his reciprocity law (Corollary 14.6.2) as an analogue of Cauchy's integral theorem [Hil32, p. 367–368]; for more on this analogy, see Vostokov [Vos2009].

Corollary 14.6.5. (Local-global principle for quaternion algebras). *Let B, B' be quaternion algebras over F. Then the following are equivalent:*

(i) $B \simeq B'$;
(ii) $\operatorname{Ram} B = \operatorname{Ram}(B')$;

(iii) $B_v \simeq B'_v$ *for all places* $v \in \operatorname{Pl} F$*; and*
(iv) $B_v \simeq B'_v$ *for all but one place* $v \in \operatorname{Pl} F$.

In particular, $B \simeq \mathrm{M}_2(F)$ *if and only if* $\operatorname{Ram} B = \emptyset$.

Proof. For the equivalence (i) ⇔ (ii) ⇔ (iii), combine Main Theorem 14.6.1 and the fact that for a noncomplex place v there is a unique division algebra over F_v. The equivalence (iii) ⇔ (iv) follows from the parity constraint, since if v is a place and $\operatorname{Ram} B \smallsetminus \{v\} = \Sigma$, then $v \in \operatorname{Ram} B$ or not according as $\#\Sigma$ is odd or even. □

Remark 14.6.6. Corollary 14.6.5 is a special case of the Albert–Brauer–Hasse–Noether theorem [AH32, BHN31]: a central simple algebra A over F such that $A_v \simeq \mathrm{M}_n(F_v)$ for all $v \in \operatorname{Pl} F$ has $A \simeq \mathrm{M}_n(F)$. See Remark 14.6.10 for further discussion.

The statement of Corollary 14.6.5 is the *local-global principle* for quaternion algebras: the isomorphism class of a quaternion algebra over a global field is determined by its isomorphism classes over the collection of local fields obtained as completions of the global field. In a similar way, we have a local-global principle for quadratic embeddings as follows.

Proposition 14.6.7. (Local-global principle for splitting/embeddings). *Let* $K \supseteq F$ *be a finite separable extension of global fields. Then the following are equivalent:*

(i) K *splits* B*, i.e.,* $B \otimes_F K \simeq \mathrm{M}_2(K)$*; and*
(ii) *For all places* $w \in \operatorname{Pl} K$*, the field* K_w *splits* B.

If $\dim_F K = 2$*, then these are further equivalent to:*

(iii) *There is an embedding* $K \hookrightarrow B$ *of* F*-algebras;*
(iv) *For all places* $v \in \operatorname{Pl} F$*, there is an embedding* $K_v \hookrightarrow B_v$ *of* F_v*-algebras; and*
(v) *Every* $v \in \operatorname{Ram} B$ *does* not *split in* K*, i.e.,* K_v *is a field for all* $v \in \operatorname{Ram} B$.

Proof. The equivalence (i) ⇔ (ii) is a consequence of Corollary 14.6.5: they are both equivalent to $\operatorname{Ram} B_K = \emptyset$, since K splits B if and only if $B \otimes_F K \simeq \mathrm{M}_2(K)$ if and only if $\operatorname{Ram}(B \otimes_F K) = \emptyset$ if and only if for all places w of K we have $B \otimes_F K_w \simeq \mathrm{M}_2(K_w)$.

The equivalence (i) ⇔ (iii) was given by Lemmas 5.4.7 and 6.4.12.

The implication (iii) ⇒ (iv) is clear. For the implication (iv) ⇒ (v), if $v \in \operatorname{Ram} B$, then B_v is a division algebra; so if K_v is not a field, then we cannot have $K_v \hookrightarrow B_v$. Finally, for (v) ⇒ (ii), let $w \in \operatorname{Pl} K$ with $w \mid v \in \operatorname{Pl} F$. If $v \notin \operatorname{Ram} B$ then already F_v splits B; otherwise, $v \in \operatorname{Ram} B$ and $K_v = K_w$ is a field with $[K_w : F_v] = 2$, so by Proposition 13.4.4, K_w splits B. □

14.6.8. The equivalences (iii) ⇔ (iv) ⇔ (v) in Proposition 14.6.7 hold also for the separable F-algebra $K = F \times F$: for there is an embedding $F \times F \hookrightarrow B$ if and only if $B \simeq \mathrm{M}_2(F)$.

We also record the statement of the Hasse–Minkowski theorem over global fields, generalizing Theorem 14.3.3.

Theorem 14.6.9. (Hasse–Minkowski). *Let F be a global field and let Q be a quadratic form over F. Then Q is isotropic over F if and only if Q_v is isotropic over F_v for all places v of F.*

Proof. The same comments as in the proof of Main Theorem 14.6.1 apply: we give a proof in section 26.8. But see also O'Meara [O'Me73, §§65–66] for a standalone class field theory proof for the case when F is a number field. □

This local-global principle for isotropy of quadratic forms is also called the **Hasse principle**. For a historical overview of the Hasse principle, and more generally Hasse's contributions in the arithmetic theory of algebras, see Fenster–Schwärmer [FS2007].

Remark 14.6.10. The fact that quaternion algebras are classified by their ramification set (Main Theorem 14.6.1) over a global field F is a consequence of the following theorem from class field theory: there is an exact sequence

$$\begin{aligned} 0 \to \operatorname{Br}(F) \to \bigoplus_v \operatorname{Br}(F_v) &\to \mathbb{Q}/\mathbb{Z} \to 0 \\ ([A_v])_v &\mapsto \textstyle\sum_v \operatorname{inv}_v [A_v] \end{aligned} \tag{14.6.11}$$

where the first map is the natural diagonal inclusion $[A] \mapsto ([A \otimes_v F_v])_v$ and the second map is the sum of the *local invariant maps* $\operatorname{inv}_v \colon \operatorname{Br}(k_v) \to \mathbb{Q}/\mathbb{Z}$ from Remark 13.4.3. The class of a quaternion algebra B in a Brauer group over a field is 2-torsion by 8.3.4, and the local invariant $\operatorname{inv}_v B_v$ is equal to $0, 1/2$ according as B_v is split or ramified, and in this way we recover the main classification theorem. (In this sense, the discriminant of a quaternion algebra captures the Brauer class of a quaternion algebra at the finite places, and the ramification set captures it fully.) The exact sequence (14.6.11) is sometimes called the **fundamental exact sequence** of global class field theory: see Milne [Milne-CFT, §VIII.4] or Neukirch–Schmidt–Wingberg [NSW2008, Theorem 8.1.17].

14.7 Theorems on norms

In the previous sections, we have seen how both local-global principles allow a nice, clean understanding of quaternion algebras—and at the same time, the norm groups play an important role in this characterization. These themes will continue through the book, so we develop them here in an important first case by describing the group $\operatorname{nrd}(B^\times) \leq F^\times$.

We retain our hypotheses that F is a global field and B a quaternion F-algebra.

14.7.1. First, we recall the calculation of the local norm groups (Lemma 13.4.9): for $v \in \operatorname{Pl} F$, we have

$$\operatorname{nrd}(B_v^\times) = \begin{cases} \mathbb{R}_{>0}^\times, & \text{if } v \in \operatorname{Ram} B \text{ is real (i.e., } B_v \simeq \mathbb{H}); \\ F_v^\times, & \text{otherwise.} \end{cases}$$

Under $B \hookrightarrow B_v$, we have $\operatorname{nrd}(B^\times) \leq \operatorname{nrd}(B_v^\times)$ for all places $v \in \operatorname{Pl} F$, and this 'places' a condition on the reduced norm at precisely the real ramified places.

14.7.2. Let $\Omega \subseteq \operatorname{Ram} B$ be the set of ramified (necessarily real) archimedean places in B. (If F is a function field, then $\Omega = \emptyset$.) Let

$$F^\times_{>_\Omega 0} := \{x \in F^\times : v(x) > 0 \text{ for all } v \in \Omega\} \tag{14.7.3}$$

be the group of elements that are positive for the embeddings $v \in \Omega$. For Ω the set of all real places, we write simply $F^\times_{>0}$ and call such elements **totally positive**.

By 14.7.1, we have $\operatorname{nrd}(B^\times) \leq F^\times_{>_\Omega 0}$. In fact, equality holds.

Main Theorem 14.7.4. (Hasse–Schilling). *We have* $\operatorname{nrd}(B^\times) = F^\times_{>_\Omega 0}$.

To prove this theorem, we will use two lemmas.

Lemma 14.7.5. *Let v be a noncomplex place of F. Let $n_v \in F_v^\times$, and if v is real suppose $n_v > 0$. Then there exists $t_v \in F_v$ such that $x^2 - t_v x + n_v$ is separable and irreducible over F_v. Moreover, if $n_v \in R_v$ then we may take $t_v \in R_v$.*

Proof. We suppose that $\operatorname{char} F_v \neq 2$ and leave the other case as an exercise (Exercise 14.22). If $n_v \notin F_v^{\times 2}$, then we can take $t_v = 0$; this includes the case where v is a real place. So suppose $n_v \in F_v^{\times 2}$. Let $K_w = F_v(\sqrt{d_v}) \supseteq F_v$ be a separable quadratic extension, and in particular $d_v \notin F_v^{\times 2}$. The quadratic form $\langle 1, -4n_v\rangle \simeq \langle 1, -1\rangle$ over F_v is a hyperbolic plane (Definition 5.4.1) so universal; let $x_v, y_v \in F_v$ be such that $x_v^2 - 4n_v y_v^2 = d_v$. We cannot have $y_v = 0$, else $x_v^2 = d_v \in F_v^{\times 2}$. Let $t_v = x_v/y_v$. Then $x^2 - t_v x + n_v$ has discriminant d_v and so is separable and irreducible.

For the second statement, multiplying by a square we may suppose without loss of generality that $d_v \in R$, so the equality $t_v^2 = d_v + 4n_v$ implies $v(t_v) \geq 0$ and $t_v \in R$. ◻

Next, we want to show that we can approximate a polynomial over a completion F_v by a polynomial over the global field F sufficiently well—the reader is invited to ignore this on a first reading and accept this intuitively as a consequence of the fact that F is dense in F_v.

Lemma 14.7.6. *Let $v \in \operatorname{Pl} F$, let $f_v(x) = x^2 - t_v x + n_v \in F_v[x]$ be a separable polynomial, and let $\epsilon > 0$. Then there exists $t, n \in F$ such that $|t - t_v|, |n - n_v| < \epsilon$ and such that $f(x) = x^2 - tx + n$ has*

$$F_v[x]/(f(x)) \simeq F_v[x]/(f_v(x)). \tag{14.7.7}$$

In particular, $f(x)$ is separable, and if $f_v(x)$ is irreducible then so is $f(x)$.

Further, if already $n_v \in F$ then we may take $n = n_v$, and similarly with t.

Proof. If F_v is nonarchimedean, the lemma follows from Corollary 13.2.9 and the fact that F is dense in F_v. The case where F_v is archimedean is straightforward: see Exercise 14.23. ◻

The same argument can be applied to several local fields at once, as follows.

Corollary 14.7.8. *Let $\Sigma \subseteq \mathrm{Pl}\, F$ be a finite set of noncomplex places. For each $v \in \Sigma$, let $f_v(x) = x^2 - t_v x + n_v \in F_v[x]$ be a separable polynomial, and let $\epsilon > 0$. Then there exists $t, n \in F$ such that for $f(x) = x^2 - tx + n$ and for all $v \in \Sigma$ we have $|t - t_v|, |n - n_v| < \epsilon$ and $F_v[x]/(f(x)) \simeq F_v[x]/(f_v(x))$. In particular, $f(x)$ is separable, and if $f_v(x)$ is irreducible for some v then so is $f(x)$.*

Further, if all $n_v = m \in F$ for $v \in \Sigma$, then we may take $n = m$, and similarly with t.

Proof. We repeat the argument of Lemma 14.7.6, using weak approximation (i.e., F is dense in $\prod_v F_v$; look ahead to Lemma 28.7.1 and the adjacent discussion) for all $v \in \Sigma$ to find t, n. □

We now conclude with a proof of the theorem on norms.

Proof of Main Theorem 14.7.4. Let $n \in F^\times_{>_\Omega 0}$. We will construct a separable quadratic extension $K \supseteq F$ with $K \hookrightarrow B$ such that $n \in \mathrm{Nm}_{K|F}(K^\times)$. To this end, by Proposition 14.6.7, it is enough to find $K \supseteq F$ such that K_v is a field for all $v \in \mathrm{Ram}\, B$.

By Lemma 14.7.5, for all $v \in \mathrm{Ram}\, B$, there exists $t_v \in F_v$ such that the polynomial $x^2 - t_v x + n \in F_v[x]$ is separable and irreducible over F_v; here if $v \in \Omega$ is real we use that $v(n) > 0$. By Corollary 14.7.8, there exists $t \in F$ such that $x^2 - tx + n$ irreducible over each F_v. Let K be the extension of F obtained by adjoining a root of this polynomial. Then K_v is a field for each ramified v, and $n \in \mathrm{Nm}_{K|F}(K^\times)$ as desired. □

Exercises

▷ 1. Complete the proof of Hilbert reciprocity (Proposition 14.2.1) in the remaining cases $(a, b) = (-1, 2), (2, 2), (-1, p), (2, p)$. In particular, show that

$$\left(\frac{-1,2}{\mathbb{Q}}\right) \simeq \left(\frac{2,2}{\mathbb{Q}}\right) \simeq \mathrm{M}_2(\mathbb{Q})$$

and

$$(a, p)_2 = (a, p)_p = \left(\frac{a}{p}\right)$$

for $a = -1, 2$ and all primes p (cf. 12.4.13).

2. Derive the law of quadratic reciprocity (14.2.4) and the supplement (14.2.5) from the statement of Hilbert reciprocity (Proposition 14.2.1).

▷ 3. Let $n \in \mathbb{Z}_{>0}$.

(a) Suppose n is of the form $n = 4^a(8b + 7)$ with $a, b \in \mathbb{Z}$. Show that there is no solution to $x^2 + y^2 + z^2 = n$ with $x, y, z \in \mathbb{Z}$. *[Hint: Look modulo 8.]*

(b) Suppose n is *not* of the form $n = 4^a(8b + 7)$ with $a, b \in \mathbb{Z}$. Show that there is a solution to $x^2 + y^2 + z^2 = n$ with $x, y, z \in \mathbb{Z}_2$. *[Hint: lift a solution modulo 8 using Hensel's lemma.]*

▷ 4. Let $n \in \mathbb{Z}$ be nonzero. Show that n is a square in $\mathbb{Q}_2$ if and only if n is of the form $n = 4^a(8b + 1)$ with $a, b \in \mathbb{Z}$.

5. Let $n > 0$ have $-n \notin \mathbb{Q}_2^{\times 2}$. Let $B = (-1, -1 \mid \mathbb{Q})$ and let $K = \mathbb{Q}(\sqrt{-n})$. Show that K splits B. *[Hint: Use the local-global principle for embeddings (Proposition 14.6.7).]* Conclude that there exists $\alpha \in B$ such that $\alpha^2 = -n$, and conclude as in Theorem 14.3.8 that n is the sum of three squares.
6. Let F be a number field. Show that every totally positive element of F is a sum of four squares of elements of F.
7. Show that the law of Hilbert reciprocity (Proposition 14.2.1) implies the law of quadratic reciprocity; with the argument given in section 14.1, this completes the equivalence of these two laws.
8. In the proof of Legendre's theorem (Theorem 14.3.4), we reduced to the case $a, b > 0$ and $c = -1$. Show that this reduction is valid.
9. In this exercise, we generalize the proof of Proposition 14.2.7 to give a more general construction of quaternion quaternion algebras. Let D be a squarefree positive integer and let $u = -1$ if D has an odd number of prime divisors, otherwise $u := 1$.

   (a) For $b \in \mathbb{Z}$ squarefree, show that $K := \mathbb{Q}(\sqrt{b})$ embeds in a quaternion algebra of discriminant D if and only if:
   - $b < 0$ if B is definite;
   - $b \not\equiv 1 \pmod 8$ if $2 \mid D$; and
   - for all odd primes $p \mid D$, we have $\left(\dfrac{b}{p}\right) \neq 1$.

   (b) Suppose b satisfies the conditions in (a) but with the further requirement that $D \mid b$, i.e., in the third condition we require $\left(\dfrac{b}{p}\right) = 0$. Let q be an odd prime such that $q^\diamond := uq$ has:
   - $\left(\dfrac{q^\diamond}{p}\right) = -1$ for all odd $p \mid D$;
   - $\left(\dfrac{q^\diamond}{p}\right) = 1$ for all odd $p \mid (b/D)$; and
   - $q^\diamond \equiv 1, 5 \pmod 8$ according as $2 \nmid D$ or $2 \mid D$.

   Note there exist infinitely many such primes q by the infinitude of primes in arithmetic progression. Then show that $B := \left(\dfrac{q^\diamond, b}{\mathbb{Q}}\right)$ has $\operatorname{disc} B = D$ and $\operatorname{Ram} B = \Sigma$.

▹ 10. Let $S \subseteq \operatorname{Pl}(\mathbb{Q})$ be eligible. For each $v \in S$, let $t_v \in \mathbb{Q}_v^\times$ be given. Show that there exists $t \in \mathbb{Q}^\times$ such that $t \in t_v \mathbb{Q}_v^{\times 2}$ for all $v \in S$ and $v_p(t) = 0$ for all $p \notin S \smallsetminus \{\infty\}$ except (possibly) for one prime $p = q$.

11. Let $F = \mathbb{Q}(\sqrt{d})$ be a real quadratic field. Find $a, b \in \mathbb{Q}^\times$ (depending on d) such that $(a, b \mid F)$ is a division ring unramified at all finite places.
12. Let $F := \mathbb{Q}(\alpha)$ where $\alpha := 2\cos(2\pi/7) = \zeta_7 + 1/\zeta_7$ with $\zeta_7 = \exp(2\pi i/7)$. Let $B := (-1, -1 \mid F)$. Compute $\operatorname{Ram}(B)$ and find a maximal order in B.

13. Let F be a global field with $\operatorname{char} F \neq 2$ and let B be a quaternion algebra over F. Let $L \supseteq F$ be a finite extension. An extension $K \supseteq F$ is **linearly disjoint** with L over F if the multiplication map $K \otimes_F L \xrightarrow{\sim} KL$ is an isomorphism of F-algebras.

    Show that there exists a splitting field $K \supseteq F$ for B such that K is linearly disjoint with L over F.

14. Show that the notion of discriminant of a quaternion algebra as the product of ramified primes is not such a great notion when R is an arbitrary Dedekind domain, as follows.

    Let $R = \mathbb{Q}[t]$; then R is a Dedekind domain. Let $F = \operatorname{Frac} R = \mathbb{Q}(t)$. Let $B_0 = (a, b \mid \mathbb{Q})$ be a division quaternion algebra ove $\mathbb{Q}$ and let $B = B_0 \otimes_{\mathbb{Q}} F = (a, b \mid \mathbb{Q}(t))$. Show that there are infinitely primes at which B is "ramified": for every prime $\mathfrak{p} = (t - c)R$, show that the algebra $B_{\mathfrak{p}}$ is a division quaternion algebra over $F_{\mathfrak{p}} \simeq \mathbb{Q}((t))$. *[Hint: See Exercise 13.6.]*

15. Using Hilbert reciprocity, one can convert the computation of an even Hilbert symbol to the computation of several odd Hilbert symbols, as follows.

    Let F be a number field, let $\mathfrak{p} \mid (2)$, and let $a, b \in F^\times$. Show that there exist (computable) $a', b' \in F^\times$ such that the following hold:

    (i) $(a, b)_{\mathfrak{p}} = (a', b')_{\mathfrak{p}}$; and
    (ii) $\operatorname{ord}_{\mathfrak{q}}(a') = \operatorname{ord}_{\mathfrak{q}}(b') = 0$ for all $\mathfrak{q} \mid (2)$ with $\mathfrak{q} \neq \mathfrak{p}$.

    Conclude that
$$(a, b)_{\mathfrak{p}} = \prod_{\substack{v \in \operatorname{Pl} F \\ v \text{ odd}}} (a', b')_v.$$

16. Let F be a number field with ring of integers $\mathbb{Z}_F$. We say an ideal $\mathfrak{b} \subseteq \mathbb{Z}_F$ is **odd** if $\operatorname{Nm}(\mathfrak{b})$ is odd, and $b \in \mathbb{Z}_F$ is **odd** if (b) is odd. For $a \in \mathbb{Z}_F$ and $\mathfrak{b} \subseteq \mathbb{Z}_F$ odd, let $\left(\dfrac{a}{\mathfrak{b}}\right)$ be the generalized Jacobi symbol, extending the generalized Legendre symbol by multiplicativity, and write $\left(\dfrac{a}{b}\right) := \left(\dfrac{a}{b\mathbb{Z}_F}\right)$ for $a, b \in \mathbb{Z}_F \smallsetminus \{0\}$ with b odd.

    (a) Let $a, b \in \mathbb{Z}_F$ satisfy $a\mathbb{Z}_F + b\mathbb{Z}_F = \mathbb{Z}_F$, with b odd, and suppose $a = a_0 a_1$ with a_1 odd. Then
$$\left(\frac{a}{b}\right)\left(\frac{b}{a_1}\right) = \prod_{v \mid 2\infty} (a, b)_v.$$

    (b) Suppose that F has a computable Euclidean function N and let $a, b \in \mathbb{Z}_F \smallsetminus \{0\}$ with b odd. Describe an algorithm using (a) to compute the Legendre symbol $\left(\dfrac{a}{b}\right)$.

17. In this exercise, we give a constructive proof of the surjectivity of the map $B \mapsto \operatorname{Ram} B$ in Main Theorem 14.6.1 in the spirit of the proof of Proposition 14.2.7 (assuming two analytic results).

Let F be a number field, and let $\Sigma \subseteq \operatorname{Pl} F$ be a finite set of noncomplex places of F of even cardinality. Let $R = \mathbb{Z}_F$ be the ring of integers of F.

(a) Let $\mathfrak{D} := \prod_{\mathfrak{p} \in \Sigma} \mathfrak{p}$ be the product of the primes corresponding to nonarchimedean places in S. Using weak approximation (see Lemma 28.7.1), show there exists $a \in \mathfrak{D}$ such that:

- $v(a) < 0$ for all real places $v \in \Sigma$ and $v(a) > 0$ for all real places $v \notin \Sigma$, if there are any; and
- $aR = \mathfrak{D}\mathfrak{b}$ with $\mathfrak{D} + \mathfrak{b} = R$ and $2R + \mathfrak{b} = R$.

In the special case where R has narrow class number 1 (that is, every ideal $\mathfrak{a} \subseteq R$ is principal $\mathfrak{a} = (a)$ and generated by an element $a \in R$ such that $v(a) > 0$ for every real places v), show that we may take $(a) = \mathfrak{D}$ and $\mathfrak{b} = R$.

(b) Show that there exists $t \in R$ coprime to $8aR$ such that the following hold:

- For all primes $\mathfrak{p} \mid \mathfrak{D}$ with $\mathfrak{p} \nmid 2R$, we have $\left(\dfrac{t}{\mathfrak{p}}\right) = -1$;
- For all primes $\mathfrak{p} \mid \mathfrak{D}$ with $\mathfrak{p} \mid 2R$, the extension $F_{\mathfrak{p}}(\sqrt{t})$ is the quadratic unramified extension of $F_{\mathfrak{p}}$, so $\left(\dfrac{t}{\mathfrak{p}}\right) = -1$ in the sense of the generalized Kronecker symbol;
- For all primes $\mathfrak{q} \mid \mathfrak{b}$ we have $\left(\dfrac{t}{\mathfrak{q}}\right) = 1$; and
- For all prime powers $\mathfrak{r}^e \parallel 8R$ with $\mathfrak{r} \nmid D$, we have $t \equiv 1 \pmod{\mathfrak{r}^e}$.

Show that t is well-defined as an element of $(R/8aR)^\times$, i.e., if $t' \equiv t \pmod{8a}$ then t' also satisfies these conditions.

(c) Using the infinitude of primes in arithmetic progression over number fields (Theorem 26.8.26), show there exists $q \in R$ a prime element (i.e., qR is a prime ideal) such that $q \equiv t \pmod{8a}$ with t as in (b) and further satisfying $v(q) < 0$ for all real places $v \in \Sigma$.

(d) Show that $B := \left(\dfrac{a, q}{F}\right)$ has $\operatorname{Ram} B = \Sigma$.

18. Let F be a global field. Show that two quaternion algebras B, B' over F are isomorphic if and only if they have the same quadratic subfields (for a quadratic extension $K \supset F$, we have $K \hookrightarrow B$ if and only if $K \hookrightarrow B'$).

[See work of Garibaldi–Saltman [GS2010] for a discussion of the fields F with $\operatorname{char} F \neq 2$ and the property that two division quaternion algebras over F with the same subfields are necessarily isomorphic. (Roughly speaking, they are the fields for which nonzero 2-torsion elements of the Brauer group can be detected using ramification.)]

19. In this exercise, we consider how ramification sets change under base extension. Let F be a global field and let $K \supseteq F$ be a finite separable extension.

(a) Let B be a quaternion algebra over F with ramification set $\operatorname{Ram} B$ and consider $B_K = B \otimes_F K$. Show that

$$\operatorname{Ram}(B_K) = \{w \in \operatorname{Pl}(K) : w \text{ lies over } v \in \operatorname{Ram} B \text{ and } 2 \nmid [K_w : F_v]\}.$$

(b) Suppose B is a division algebra and $[K : F]$ is odd. Show that B_K is a division algebra.
(c) As a converse to (a), suppose that $\Sigma_K \subseteq \mathrm{Pl}(K)$ is a finite subset of noncomplex places of K of even cardinality with the property that if $w \in \Sigma_K$ lies over $v \in \mathrm{Pl}\, F$, then $[K_w : F_v]$ is odd and moreover

$$\{w \in \mathrm{Pl}\, F : w \text{ lies over } v \text{ and } [K_w : F_v] \text{ is odd}\} \subseteq \Sigma_K.$$

Show that there exists a quaternion algebra B over F with the property that $\mathrm{Ram}(B_K) = \Sigma_K$. (We say that the quaternion algebra associated to the set Σ_K **descends** to F.)
(d) As a special case, what do (a) and (c) say when $[K : F] = 2$?
(e) Restate (a) and (b) in terms of the kernel of the map $\mathrm{Br}(F)[2] \to \mathrm{Br}(K)[2]$ induced by $[B] \mapsto [B_K]$ (see Remark 14.6.10).

20. Let R be a global ring with $F = \mathrm{Frac}\, R$, and let $K \supseteq F$ be a finite Galois extension with S the integral closure of R in K. Let B be a quaternion algebra over F and consider $B_K = B \otimes_F K$. Then $\mathrm{Gal}(K \,|\, F)$ acts naturally on B_K via $\sigma(\alpha \otimes x) = \alpha \otimes \sigma(x)$. (This action is not by K-algebra isomorphism!)

Show that there exists a maximal S-order $O \subseteq B_K$ stable under $\mathrm{Gal}(K \,|\, F)$, i.e., $\sigma(O) = O$ for all $\sigma \in \mathrm{Gal}(K \,|\, F)$.

▹ 21. Let F be a global field, let $v_1, \dots, v_r$ be places of F, and for each v_i suppose we are given the condition *ramified*, *split*, or *inert*. Show that there exists a separable quadratic extension $K \supseteq F$ that K_{v_i} satisfies the given condition for each i. *[Hint: follow the proof of Main Theorem 14.7.4.]*

22. Let F be a global field, let $B_1, B_2, \dots, B_r$ be quaternion algebras over F, and let $B := B_1 \otimes B_2 \otimes \cdots \otimes B_r$. Recalling section 8.2, show (in as many ways as you can) that $B \simeq \mathrm{M}_{2^{r-1}}(B')$ for a quaternion algebra B' over F. (Recalling 8.3, by Merkurjev's theorem this shows the class of every element in the 2-torsion of the Brauer group $\mathrm{Br}(F)[2]$ is represented by a quaternion algebra.)

▹ 23. Let F_v be a local field with $\mathrm{char}\, F_v = 2$. Let $n \in F_v$. Show that there exists $t \in F_v$ such that $x^2 - tx + n$ is separable and irreducible.

▹ 24. Prove Lemma 14.7.6 for v an archimedean place.

25. In this advanced exercise following up on Exercise 9.8, we consider features of quaternion algebras and orders in the case of a global *function* field, assuming background in algebraic geometry.

Let X be a smooth, projective, geometrically integral curve over a finite field k; then X is a separated, integral Dedekind scheme. Let $\mathscr{O}_X$ be its structure sheaf. Let F be its function field, and let B be a quaternion algebra over F. Define a sheaf of **$\mathscr{O}_X$-orders** in B, or simply an **$\mathscr{O}_X$-order** in B, to be an $\mathscr{O}_X$-lattice $\mathscr{B}$ in B such that for each open set $U \subseteq X$, the $\mathscr{O}_X(U)$-lattice $\mathscr{B}(U)$ is a subring of B. We recall the local-global dictionary for $\mathscr{O}_X$-lattices (Exercise 9.16(c)).

In parts (a)–(c) we work out an example: let $X = \mathbb{P}^1$ with function field $F = k(t)$ where $\operatorname{div} t = (0) - (\infty)$. Suppose that $\operatorname{char} k \neq 2$, and let $u \in k^\times \smallsetminus k^{\times 2}$. Let B be the quaternion algebra with $\operatorname{Ram}(B) = \{(0), (\infty)\}$.

(a) Show that $B = (u, t \mid F)$.
(b) Let $U = \operatorname{Spec} k[t] = X \smallsetminus \{\infty\}$. Show that there exists a unique $\mathscr{O}_X$-order $\mathscr{B}$ in B with $\mathscr{B}(U) = k[t] + k[t]i + k[t]j + k[t]ij$ and stalk $\mathscr{B}_{(\infty)}$ a maximal $\mathscr{O}_{X,(\infty)}$-order. Describe explicitly $\mathscr{B}_{(\infty)}$ and $\mathscr{B}(\operatorname{Spec} k[1/t])$ as orders in B.
(c) With $\mathscr{B}$ from (b), show that $\mathscr{B}(X) = k[i]$.

Restoring generality, let $\mathscr{B}$ be an $\mathscr{O}_X$-order such that $\mathscr{B}(U)$ is a maximal $\mathscr{O}_X(U)$-order in B for all affine open sets U.

(d) Show that $\mathscr{B}(X)$ has a zero divisor if and only if $\mathscr{B}(X) \simeq \mathrm{M}_2(k)$ if and only if $B \simeq \mathrm{M}_2(F)$.
(e) Show that $\mathscr{B}(X)$ is a k-algebra with a nondegenerate standard involution.
(f) Suppose that B is a division algebra. Show that either $\mathscr{B}(X) = k$ or $\mathscr{B}(X) \simeq k_2$ is the quadratic extension of k.
(g) Still supposing that B is a division algebra, show that if $\mathscr{B}(X) = k_2$, then every ramified place of B has odd degree. *[Hint: show that $B \simeq (K, b \mid F)$ where $K = Fk_2$ is the constant field extension of F of degree 2, and $b \in F^\times \smallsetminus k^\times$. Compute the Hilbert symbol at $v \in \operatorname{Ram}(B)$ to show $v(b)$ is odd.]*

This chapter is licensed under the terms of the Creative Commons Attribution-NonCommercial 4.0 International License (http://creativecommons.org/licenses/by-nc/4.0/), which permits any noncommercial use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license and indicate if changes were made.

The images or other third party material in this chapter are included in the chapter's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the chapter's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.

Chapter 15

Discriminants

Discriminants measure volume and arithmetic complexity, and they simultaneously encode ramification. We devote this chapter to their study.

15.1 ▹ Discriminantal notions

Let $x_1, \dots, x_n \in \mathbb{R}^n$, and let A be the matrix with columns x_i. Then the parallelopiped with edges from the origin to x_i has volume $|\det(A)|$. We can compute this volume in another way:

$$\det(A)^2 = \det(A^{\mathsf{t}}A) = \det(M) \tag{15.1.1}$$

where M has ijth entry equal to the ordinary dot product $x_i \cdot x_j$.

The absolute discriminant of a number field is a volume and a measure of arithmetic complexity, as follows. If $x_1, \dots, x_n$ is a $\mathbb{Z}$-basis for $\mathbb{Z}_F$ and $\iota : F \hookrightarrow F \otimes_{\mathbb{Q}} \mathbb{R} \simeq \mathbb{R}^n$ (normalized with an extra factor of $\sqrt{2}$ at the complex places), then the volume of $\mathbb{Z}_F$ in this embedding is the absolute determinant of the matrix with columns $\iota(x_i)$, and its square is defined to be the absolute discriminant of F. Replacing the dot product in the definition of M in (15.1.1) with the trace form $(x, y) \mapsto \operatorname{Tr}_{F/\mathbb{Q}}(xy)$, we see that the absolute discriminant is a positive integer. A prime p is ramified in F if and only if it divides the discriminant, so this volume also records arithmetic properties of F.

More generally, whenever we have a symmetric bilinear form $T\colon V \times V \to F$ on a finite-dimensional F-vector space V, there is a volume defined by the determinant $\det(T(x_i, x_j))_{i,j}$: and when T arises from a quadratic form Q, this is volume is the discriminant of Q (up to a normalizing factor of 2 in odd degree, see 6.3.1). In particular, if B is a finite-dimensional algebra over F, there is a bilinear form

$$\begin{aligned} B \times B &\to F \\ (\alpha, \beta) &\mapsto \operatorname{Tr}_{B|F}(\alpha\beta) \end{aligned}$$

(or, when B is semisimple, the bilinear form associated to the reduced trace trd) and so we obtain a discriminant—a “squared” volume—measuring in some way the complexity of B. As in the commutative case, discriminants encode ramification.

In this chapter, we establish basic facts about discriminants, including how they behave under inclusion (measuring index) and localization. To illustrate, let B be a

© The Author(s) 2021

J. Voight, *Quaternion Algebras*, Graduate Texts in Mathematics 288,
https://doi.org/10.1007/978-3-030-56694-4_15

quaternion algebra over $\mathbb{Q}$ and let $O \subset B$ be an order. We define the **discriminant** of O to be

$$\operatorname{disc}(O) := |\det(\operatorname{trd}(\alpha_i\alpha_j))_{i,j}| \in \mathbb{Z}_{>0} \tag{15.1.2}$$

where $\alpha_1, \dots, \alpha_4$ is a $\mathbb{Z}$-basis for O. For example, if $B = \left(\dfrac{a,b}{\mathbb{Q}}\right)$ with $a, b \in \mathbb{Z} \smallsetminus \{0\}$, then the standard order $O = \mathbb{Z} + \mathbb{Z}i + \mathbb{Z}j + \mathbb{Z}k$ has

$$\operatorname{disc}(O) = (4ab)^2;$$

indeed, this is the discriminant of the quadratic form $\langle 1, -a, -b, ab\rangle$, the reduced norm restricted to O. If $a, b < 0$, i.e. B is definite, then the reduced norm is a Euclidean norm on $B_\infty = B \otimes_{\mathbb{Q}} \mathbb{R} \simeq \mathbb{H}$; normalizing with an extra factor $\sqrt{2}$, the discriminant is square of the covolume of the lattice $O \subset B_\infty$. For example, the Lipschitz order $\mathbb{Z}\langle i, j\rangle$ (11.1.1) has $\operatorname{disc}(\mathbb{Z}\langle i, j\rangle) = 4^2$, the square of the covolume of the lattice $(\sqrt{2}\mathbb{Z})^4 \subseteq \mathbb{R}^4$.

If $O' \supseteq O$, then $\operatorname{disc}(O) = [O' : O]^2 \operatorname{disc}(O')$; in particular $O' = O$ if and only if $\operatorname{disc}(O') = \operatorname{disc}(O)$. It follows that the discriminant of an order is always a square, so we define the **reduced discriminant** $\operatorname{discrd}(O)$ to be the positive integer square root, and $\operatorname{discrd}(O)^2 = \operatorname{disc}(O)$. The discriminant of an order measures how far the order is from being a maximal order. We will show (Theorem 15.5.5) that O is a maximal order if and only if $\operatorname{discrd}(O) = \operatorname{disc} B$, where $\operatorname{disc} B$ is the (squarefree) product of primes ramified in B.

In an extension of Dedekind domains, the different of the extension is an ideal whose norm is the discriminant of the extension (see Neukirch [Neu99, §III.2]). The different is perhaps not as popular as its discriminant cousin, but it has many nice properties, including easy-to-understand behavior under base extension. Similar conclusions holds in the noncommutative context (presented in section 15.6).

15.2 Discriminant

For further reference on discriminants, see Reiner [Rei2003, §10, §14].

Let R be a noetherian domain and let $F = \operatorname{Frac} R$. Let B be a semisimple algebra over F with $\dim_F B = n$. For elements $\alpha_1, \dots, \alpha_n \in B$, we define

$$d(\alpha_1, \dots, \alpha_n) := \det(\operatorname{trd}(\alpha_i\alpha_j))_{i,j=1,\dots,n}. \tag{15.2.1}$$

Let $I \subseteq B$ be an R-lattice.

Definition 15.2.2. The **discriminant** of I is the R-submodule $\operatorname{disc}(I) \subseteq F$ generated by the set

$$\{d(\alpha_1, \dots, \alpha_n) : \alpha_1, \dots, \alpha_n \in I\}.$$

15.2.3. If $I = O$, then for $\alpha_1, \dots, \alpha_n \in O$ we have $\alpha_i\alpha_j \in O$ and so $\operatorname{trd}(\alpha_i\alpha_j) \in R$ for all i, j. Thus $d(\alpha_1, \dots, \alpha_n) \in R$ and therefore $\operatorname{disc}(O) \subseteq R$.

Remark 15.2.4. When working over $\mathbb{Z}$, it is common to take the discriminant instead to be the positive generator of the discriminant as an ideal; passing between these should cause no confusion.

Although Definition 15.2.2 may look unwieldly, it works as well in the commutative case as in the noncommutative case. Right away, we see that if $O \subseteq O'$ are R-orders, then $\operatorname{disc}(O') \mid \operatorname{disc}(O)$.

The function d itself transforms in a nice way under a change of basis, as follows.

Lemma 15.2.5. *Let $\alpha_1, \dots, \alpha_n \in B$ and suppose $\beta_1, \dots, \beta_n \in B$ are of the form $\beta_i = \sum_{j=1}^n m_{ij}\alpha_j$ with $m_{ij} \in F$. Let $M = (m_{ij})_{i,j=1,\dots,n}$. Then*

$$d(\beta_1, \dots, \beta_n) = \det(M)^2 d(\alpha_1, \dots, \alpha_n). \tag{15.2.6}$$

Proof. By properties of determinants, if $\beta_1, \dots, \beta_n$ are linearly dependent (over F) then $d(\beta_1, \dots, \beta_n) = 0$ and either $\alpha_1, \dots, \alpha_n$ are also linearly dependent or $\det(M) = 0$, and in either case the equality (15.2.6) holds trivially.

So suppose that $\beta_1, \dots, \beta_n$ are linearly independent, then $\alpha_1, \dots, \alpha_n$ are also linearly independent and the matrix M, a change of basis matrix, is invertible. By Gaussian reduction, we can write M as a product of elementary matrices (a matrix that coincides with the identity matrix except for a single off-diagonal entry), permutation matrices (a matrix interchanging rows suffices), and a diagonal matrix; it is enough to check that the equality holds when M is a matrix of one of these forms. And for such a matrix, the equality can be checked in a straightforward manner using the corresponding property of determinants. □

Corollary 15.2.7. *If I is free as an R-module, and $\alpha_1, \dots, \alpha_n$ is an R-basis for I, then*

$$\operatorname{disc}(I) = d(\alpha_1, \dots, \alpha_n)R.$$

Proof. The matrix M writing any other $\beta_1, \dots, \beta_n \in I$ in terms of the basis has $M \in \mathrm{M}_n(R)$ so $\det(M) \in R$, and therefore $d(\beta_1, \dots, \beta_n) \in d(\alpha_1, \dots, \alpha_n)R$ by Lemma 15.2.5. □

15.2.8. More generally, if I is completely decomposable with

$$I = \mathfrak{a}_1\alpha_1 \oplus \cdots \oplus \mathfrak{a}_n\alpha_n$$

such as in (9.3.7), then from (15.2.6)

$$\operatorname{disc}(I) = (\mathfrak{a}_1 \cdots \mathfrak{a}_n)^2 d(\alpha_1, \dots, \alpha_n).$$

More generally, the discriminant is well-behaved under automorphisms because the reduced trace is so.

Corollary 15.2.9. *If $\phi \colon B \xrightarrow{\sim} B$ is an F-algebra automorphism, then* $\operatorname{disc}(\phi(I)) = \operatorname{disc}(I)$.

Proof. By Proposition 7.8.6, we have $\operatorname{trd}(\phi(\alpha\beta)) = \operatorname{trd}(\alpha\beta)$ for all $\alpha, \beta \in B$. Therefore, for all $\alpha_1, \dots, \alpha_n \in B$ we have $d(\phi(\alpha_1), \dots, \phi(\alpha_n)) = d(\alpha_1, \dots, \alpha_n)$; the result $\operatorname{disc}(\phi(I)) = \operatorname{disc}(I)$ follows. □

Our primary interest will be in the case $I = O$.

Example 15.2.10. Suppose $\operatorname{char} F \neq 2$. Let $B := (a, b \mid F)$ with $a, b \in R$. Let $O := R \oplus Ri \oplus Rj \oplus Rij$ be the standard order. Then $\operatorname{disc}(O)$ is the principal R-ideal generated by

$$d(1, i, j, ij) = \det \begin{pmatrix} 2 & 0 & 0 & 0 \\ 0 & 2a & 0 & 0 \\ 0 & 0 & 2b & 0 \\ 0 & 0 & 0 & -2ab \end{pmatrix} = -(4ab)^2.$$

The calculation when $\operatorname{char} F = 2$ is requested in Exercise 15.1.

Example 15.2.11. Let $B := \mathrm{M}_n(F)$ and $O := \mathrm{M}_n(R)$. Then $\operatorname{disc}(O) = R$ (Exercise 15.2).

15.2.12. Let $B := (K, b \mid F)$ be a quaternion algebra over F with $b \in R$ and let S be an R-order in K. Let $O := S \oplus Sj$; then O is an R-order in B by Exercise 10.7. We have $\operatorname{disc}(O) = b^2 \operatorname{disc}(S)^2$, by Exercise 15.4.

In particular, let F be a nonarchimedean local field, let R be its valuation ring and $\mathfrak{p} = R\pi$ its maximal ideal, and let B be a division quaternion algebra over F. Then by Theorem 13.3.11, we have $B \simeq (K, \pi \mid F)$ with $K \supseteq F$ an unramified separable quadratic extension of F. The valuation ring S of K has $\operatorname{disc}(S) = R$, so the valuation ring $O = S \oplus Sj$ of B has discriminant $\operatorname{disc}(O) = \mathfrak{p}^2$.

15.2.13. Equation (15.2.6) and the fact that $I_{(\mathfrak{p})} = I \otimes_R R_{(\mathfrak{p})}$ implies the equality

$$\operatorname{disc}(I_{(\mathfrak{p})}) = \operatorname{disc}(I)_{(\mathfrak{p})}$$

on localizations and for the same reason an equality for the completions $\operatorname{disc}(I_{\mathfrak{p}}) = \operatorname{disc}(I)_{\mathfrak{p}}$. In other words, the discriminant respects localization and completion and can be computed locally. Therefore, by the local-global principle (Lemma 9.4.6),

$$\operatorname{disc}(I) = \bigcap_{\mathfrak{p}} \operatorname{disc}(I_{(\mathfrak{p})}).$$

Lemma 15.2.14. *If B is separable as an F-algebra and I is projective as an R-module, then* $\operatorname{disc}(I)$ *is a nonzero projective fractional ideal of R.*

Proof. Since I is an R-lattice, there exist elements $\alpha_1, \dots, \alpha_n$ which are linearly independent over F. Since B is separable, by Theorem 7.9.4, trd is a nondegenerate bilinear pairing on B so $\operatorname{disc}(I)$ is a nonzero ideal of R. It follows from Lemma 15.2.5 that $\operatorname{disc}(I)$ is finitely generated as an R-module, since this is true of I: we apply d to all subsets of a set of generators for I as an R-module. To show that $\operatorname{disc}(I)$ is projective, by 9.2.1 we show that $\operatorname{disc}(I)$ is locally principal. Let $\mathfrak{p}$ be a prime ideal of R. Since I is a projective R-module, its localization $I_{(\mathfrak{p})}$ is free; thus from Corollary 15.2.7, we conclude that $\operatorname{disc}(I)_{(\mathfrak{p})} = \operatorname{disc}(I_{(\mathfrak{p})})$ is principal over $R_{(\mathfrak{p})}$ and generated by $\operatorname{disc}(\alpha_1, \dots, \alpha_n)$ for an $R_{(\mathfrak{p})}$-basis $\alpha_1, \dots, \alpha_n$ of I, as desired. □

We conclude this section comparing lattices by their index and discriminant as follows. We recall the definition of index (section 9.6).

Lemma 15.2.15. *Let $I, J \subseteq B$ be projective R-lattices. Then*

$$\mathrm{disc}(I) = [J : I]_R^2 \,\mathrm{disc}(J).$$

Moreover, if $I \subseteq J$, then $\mathrm{disc}(I) = \mathrm{disc}(J)$ *if and only if $I = J$.*

Proof. For the first statement, we argue locally, and combine (15.2.6) and Lemma 9.6.4. For the second statement, clearly $\mathrm{disc}(J) \subseteq \mathrm{disc}(I)$, and if $I = J$ then equality holds; and conversely, from $\mathrm{disc}(I) = [J : I]_R^2 \,\mathrm{disc}(J) = \mathrm{disc}(J)$ we conclude $[J : I]_R = R$, hence $J = I$ by Proposition 9.6.8. □

Remark 15.2.16. We defined the discriminant for semisimple algebras so that it is given in terms of the reduced trace. This definition extends to an arbitrary finite-dimensional F-algebra B, replacing the reduced trace by the algebra trace $\mathrm{Tr}_{B|F}$. If B is a central simple F-algebra of dimension n^2, then $n\,\mathrm{trd} = \mathrm{Tr}_{B|F}$ so when $n \in F^\times$ one can recover the discriminant as we have defined it here from the more general definition; but if $n = 0 \in F$ then the discriminant of B computed with the algebra trace will be zero.

15.3 Quadratic forms

Essentially the same definition of discriminant (Definition 15.2.2) applies to quadratic modules, as follows. We recall 6.3.1, where the discriminant was defined in all characteristics.

Let $Q\colon M \to L$ be a quadratic module over R (Definition 9.7.3) with $\mathrm{rk}\, M = n$ and associated bilinear map $T\colon M \times M \to L$.

15.3.1. Let $x_1, \dots, x_n \in M$ and $f \in L^\vee := \mathrm{Hom}_R(L, R)$. If n is even, we define

$$d(x_1, \dots, x_n; f) := \det(f(T(x_i, x_j)))_{i,j=1,\dots,n}. \tag{15.3.2}$$

If n is odd, then by specializing the universal determinant as in 6.3.4, we define

$$d(x_1, \dots, x_n; f) := (\det/2)(f(T(x_i, x_j)))_{i,j=1,\dots,n}. \tag{15.3.3}$$

The **discriminant** of Q is then the ideal $\mathrm{disc}(Q) \subseteq R$ generated by the set

$$\{d(x_1, \dots, x_n; f) : x_1, \dots, x_n \in M,\ f \in L^\vee\}. \tag{15.3.4}$$

15.3.5. If M, L are free with R-basis $x_1, \dots, x_n$ and e, respectively, then letting $f \in L^\vee$ the dual to e with $f(e) = 1$ gives

$$\mathrm{disc}(Q) = d(x_1, \dots, x_n; f)R.$$

In particular, since M, L are projective and therefore locally free over R, the discriminant of Q is locally free and hence a projective R-ideal.

Lemma 15.3.6. *The discriminant of a quadratic module is well-defined up to similarity.*

Proof. Let $Q\colon M \to L$ and $Q'\colon M' \to L'$ be quadratic modules over R similar by $g : M \xrightarrow{\sim} M'$ and $h : L \xrightarrow{\sim} L'$. It suffices to check the invariance locally, so to this end we may suppose that the modules are free; choose a basis $M = \sum_{i=1}^n Rx_i$ and $L = Re$, and let $x_i' = g(x_i)$ and $e' = h(e)$. Then $M' = \sum_{i=1}^n Rx_i'$ and $L' = Re'$. Let f, f' be dual to e, e'; then postcomposing Q and Q' by f, f' we may suppose $L = L' = R$ and h is the identity.

We then have $Q'(g(x)) = Q(x)$ for all $x \in M$, so the same is true of the associated bilinear forms T, T'. But then $d(x_1', \dots, x_n') = d(x_1, \dots, x_n)$, and by 15.3.5 this implies $\operatorname{disc}(Q) = \operatorname{disc}(Q')$ as ideals of R. □

15.3.7. Let B be a finite-dimensional F-algebra with a standard involution. Then the reduced norm is a quadratic form on B with associated bilinear form $T(\alpha, \beta) = \operatorname{trd}(\alpha\overline{\beta})$. Although the bilinear form differs by the presence of this standard involution from the definition of discriminant in (15.2.1), the resulting discriminants are the same (up to $R^\times$): see Exercise 15.13.

Lemma 15.3.8. *The quadratic module Q is nonsingular if and only if* $\operatorname{disc}(Q) = L$.

In particular, suppose that $M \simeq R^n$ is free with basis e_i and $L = R$, and let $[T] := (T(e_i, e_j))_{i,j} \in \mathrm{M}_n(R)$ be the Gram matrix in this basis. Then Q is nonsingular if and only if $\det([T]), (\det/2)([T]) \in R^\times$ *according as n is even or odd.*

Proof. The map $T\colon M \to \operatorname{Hom}_R(M, L)$ is an isomorphism if and only if it is an isomorphism in every localization, so we may suppose that Q is free, with $M = R^n$ and $L = R$, which is to say we may prove the second statement in the case where R is local, with maximal ideal $\mathfrak{p}$ and residue field $k := R/\mathfrak{p}$. Let $Q \bmod \mathfrak{p}\colon M \otimes_R k \to k$ be the reduction of Q; its Gram matrix is $[T] \bmod \mathfrak{p} \in \mathrm{M}_n(k)$. Over the field k, we have that $Q \bmod \mathfrak{p}$ is nonsingular if and only if it is nondegenerate if and only if $\det[T], (\det/2)([T]) \neq 0$ according as n is even or odd; since R is local, these are equivalent to asking that these values are in $R^\times$. An application of Nakayama's lemma then implies the result. □

15.4 Reduced discriminant

In this section, we extract a square root of the discriminant for quaternion orders. Indeed, in Example 15.2.10, we saw that the discriminant of the standard R-order $O \subseteq B = (a, b \mid F)$ is $\operatorname{disc}(O) = (4ab)^2R$, a square. If O' is another projective R-order, then $\operatorname{disc}(O') = [O : O']_R^2 \operatorname{disc}(O)$ by Lemma 15.2.15, so in fact the discriminant of every R-order is the square of an R-ideal.

In fact, there is a way to define this square root directly, inspired by vector calculus.

15.4.1. If $u, v, w \in \mathbb{R}^3$ then $|u\cdot(v\times w)|$, the absolute value of the so-called **mixed product** (or **scalar triple product** or **box product**), is the volume of the parallelopiped defined by u, v, w; identifying $\mathbb{R}^3 \simeq \mathbb{H}^0$ as in section 2.4, from (2.4.10) we can write

$$2u \cdot (v \times w) = u \cdot (vw - wv) = -\operatorname{trd}(u(vw - wv)).$$

For example, $2 = -2i \cdot (j \times k) = -i \cdot (jk - kj) = \operatorname{trd}(ijk)$.

More generally (and carefully attending to the factors of 2) we make the following definition. Let B be a quaternion algebra over F.

15.4.2. For $\alpha_1, \alpha_2, \alpha_3 \in B$, we define

$$\begin{aligned} m(\alpha_1, \alpha_2, \alpha_3) &:= \operatorname{trd}((\alpha_1\alpha_2 - \alpha_2\alpha_1)\overline{\alpha_3}) \\ &= \alpha_1\alpha_2\overline{\alpha_3} - \alpha_2\alpha_1\overline{\alpha_3} - \alpha_3\overline{\alpha_2}\,\overline{\alpha_1} + \alpha_3\overline{\alpha_1}\,\overline{\alpha_2}. \end{aligned}$$

Lemma 15.4.3. *The form $m \colon B \times B \times B \to F$ is an alternating trilinear form which is well-defined as a form on B/F.*

Proof. The form is alternating because for all $\alpha_1, \alpha_2 \in B$ we have $m(\alpha_1, \alpha_1, \alpha_2) = 0$ and

$$m(\alpha_1, \alpha_2, \alpha_1) = \operatorname{trd}((\alpha_1\alpha_2 - \alpha_2\alpha_1)\overline{\alpha_1}) = \operatorname{trd}(\operatorname{nrd}(\alpha_1)\alpha_2) - \operatorname{trd}(\alpha_2 \operatorname{nrd}(\alpha_1)) = 0$$

and similarly $m(\alpha_1, \alpha_2, \alpha_2) = 0$. The trilinearity follows from the linearity of the reduced trace. Finally, from these two properties, the descent to B/F follows from the computation $m(1, \alpha_1, \alpha_2) = 0$ for all $\alpha_1, \alpha_2 \in B$.

(Alternatively, one can check that the pairing descends to B/F first, so that the involution becomes $\overline{\alpha + F} = -\alpha + F$, and then the alternating condition is immediate.) □

Definition 15.4.4. Let $I \subseteq B$ be an R-lattice. The **reduced discriminant** of I is the R-submodule $\operatorname{discrd}(I)$ of F generated by

$$\{m(\alpha_1, \alpha_2, \alpha_3) : \alpha_1, \alpha_2, \alpha_3 \in I\}.$$

15.4.5. If $\alpha_i, \beta_i \in B$ with $\beta_i = M\alpha_i$ for some $M \in \mathrm{M}_3(F)$, then

$$m(\beta_1, \beta_2, \beta_3) = \det(M) m(\alpha_1, \alpha_2, \alpha_3) \tag{15.4.6}$$

by Exercise 15.10. It follows that if $I \subseteq J$ are projective R-lattices in B, then

$$\operatorname{discrd}(I) = [J : I] \operatorname{discrd}(J).$$

Lemma 15.4.7. *If I is a projective R-lattice in B, then* $\operatorname{disc}(I) = \operatorname{discrd}(I)^2$.

Proof. First, we claim that

$$m(i, j, ij)^2 = -d(1, i, j, ij).$$

If $\operatorname{char} F \neq 2$, then $\operatorname{disc}(1, i, j, ij) = -(4ab)^2$ by Example 15.2.10 and

$$m(i, j, ij) = \operatorname{trd}((ij - ji)\overline{ij}) = \operatorname{trd}(2ij(\overline{ij})) = 4ab,$$

as claimed. See Exercise 15.1 for the case $\operatorname{char} F = 2$. This computation verifies the result for the order $O = R \oplus Ri \oplus Rj \oplus Rij$.

The lemma now follows using (15.2.6) and (15.4.6), for it shows that

$$m(\alpha_1, \alpha_2, \alpha_3)^2 = -d(1, \alpha_1, \alpha_2, \alpha_3)$$

for all $\alpha_1, \alpha_2, \alpha_3 \in B$, and the latter generate $\operatorname{discrd}(I)$ by Exercise 15.7. □

The notions in this section extend more generally to an arbitrary algebra B with a standard involution.

15.5 Maximal orders and discriminants

We now relate discriminants to maximal orders. Throughout this section, we suppose that R is a Dedekind domain. We record the following important principle.

Lemma 15.5.1. *Let $O \subseteq O'$ be R-orders. Then $O = O'$ if and only if* $\operatorname{disc} O = \operatorname{disc} O'$.

Proof. In the nontrivial direction, by Lemma 15.2.15 we have

$$\operatorname{disc} O = [O' : O]_R^2 \operatorname{disc}(O')$$

so $\operatorname{disc} O = \operatorname{disc} O'$ if and only if $O = O'$. □

First, we ensure the existence of maximal orders (cf. 10.4.2) using the discriminant.

Proposition 15.5.2. *There exists a maximal R-order $O \subseteq B$, and every order O is contained in a maximal R-order $O' \subseteq B$.*

Proof. The algebra B has at least one R-order O as the left- or right-order of a lattice 10.2.5. If O is not maximal, then there exists an order $O' \supsetneq O$ with $\operatorname{disc}(O') \supsetneq \operatorname{disc}(O)$ by Lemma 15.5.1. If O' is maximal, we are done; otherwise, we can continue in this way to obtain orders $O = O_1 \subsetneq O_2 \subsetneq \dots$ and an ascending chain of ideals $\operatorname{disc}(O_1) \subsetneq \operatorname{disc}(O_2) \subsetneq \dots$ of R; but since R is noetherian, the latter stabilizes after finitely many steps, and the resulting order is then maximal, by Lemma 15.2.15. □

Using the discriminant as a measure of index, we can similarly detect when orders are maximal. We recall (10.4.3) that the property of being maximal is a local property, so we begin with the local matrix case.

Lemma 15.5.3. *Suppose that R is a DVR, and let $O \subseteq B := \mathrm{M}_n(F)$ be an R-order. Then O is maximal if and only if* $\operatorname{disc} O = R$.

Proof. First, suppose O is maximal. Then by Corollary 10.5.5, we conclude $O \simeq \mathrm{M}_n(R)$ (conjugate in B). By Corollary 15.2.9, we have $\operatorname{disc} O = \operatorname{disc} \mathrm{M}_n(R)$; we computed in Example 15.2.11 that $\operatorname{disc} \mathrm{M}_n(R) = R$, as claimed. The converse follows by taking O' a maximal order containing O (furnished by Proposition 15.5.2) and applying Lemma 15.5.1. □

Example 15.5.4. By 15.2.12, if F is a nonarchimedean local field with valuation ring R and B is a division quaternion algebra over F, then the valuation ring $O \subset B$ is the unique maximal order (Theorem 13.3.11) with $\operatorname{disc} O = \mathfrak{p}^2$ and $\operatorname{discrd} O = \mathfrak{p}$. Arguing as in Lemma 15.5.3, we find that an R-order in B is maximal if and only if it has reduced discriminant $\mathfrak{p}$.

Maximality can be detected over global rings in terms of discriminants, as follows.

Theorem 15.5.5. *Let R be a global ring with field of fractions F, let B be a quaternion algebra over F, and let $O \subseteq B$ be an R-order. Then O is maximal if and only if*

$$\operatorname{discrd}(O) = \operatorname{disc}_R(B). \tag{15.5.6}$$

Proof. Suppose that O is maximal. Then $O_\mathfrak{p}$ is maximal for all primes $\mathfrak{p}$ of R. If $B_\mathfrak{p} \simeq \mathrm{M}_2(F_\mathfrak{p})$ is split, then by Lemma 15.5.3, $\operatorname{discrd} O_\mathfrak{p} = R_\mathfrak{p}$; if $B_\mathfrak{p}$ is a division algebra, then $\operatorname{discrd} O_\mathfrak{p} = \mathfrak{p}R_\mathfrak{p}$. Since discriminants are defined locally, we conclude that

$$\operatorname{discrd}(O) = \prod_{\mathfrak{p} \in \operatorname{Ram} B \smallsetminus S} \mathfrak{p} = \operatorname{disc}_R(B)$$

if R as a global ring is the ring of S-integers.

In the other direction, if (15.5.6) holds, we choose $O' \supseteq O$ be a maximal R-superorder and conclude that $\operatorname{disc}(O) = \operatorname{disc}_R(B)^2 = \operatorname{disc}(O')$ so $O = O'$ is maximal by Lemma 15.5.1. □

Example 15.5.7. We recall Example 14.2.13, giving an explicit description of quaternion algebras $B = (a, b \mid \mathbb{Q})$ of prime discriminant $D = p$. We now exhibit an explicit maximal order in each of these algebras.

For $p = 2$, we have $B = (-1, -1 \mid \mathbb{Q})$ and take $O \subseteq B$ the Hurwitz order.

For $p \equiv 3 \pmod 4$, we took $B = (-p, -1 \mid \mathbb{Q})$. The order $O := \mathbb{Z}\langle (1+i)/2, j\rangle = S \oplus Sj$ with $S := \mathbb{Z}[(1 + i)/2]$ has $\operatorname{discrd} O = p$ by 15.2.12, so O is maximal by Theorem 15.5.5.

For $p \equiv 1 \pmod 4$, we had $B = (-p, -q \mid \mathbb{Q})$ where $q \equiv 3 \pmod 4$ is prime and $\left(\dfrac{q}{p}\right) = -1$, so that by qudaratic reciprocity $-pq = -qp = 1$. In this case, let $c \in \mathbb{Z}$ be such that $c^2 \equiv -p \pmod q$. Then

$$O := \mathbb{Z} \oplus \mathbb{Z}\frac{1 + j}{2} \oplus \mathbb{Z}\frac{i(1 + j)}{2} \oplus \mathbb{Z}\frac{(c + i)j}{q}$$

is a maximal order: one checks that O is closed under multiplication (in particular, the basis elements are integral), and then that $\operatorname{disc} O = p$. The order $\mathbb{Z}\langle i, (1+j)/2\rangle \subseteq O$ has the larger reduced discriminant pq, hence the need for a denominator q in the fourth element.

For further discussion of explicit maximal orders over $\mathbb{Z}$, see Ibukiyama [Ibu82, pp. 181–182] or Pizer [Piz80a, Proposition 5.2]. For a more general construction, see Exercise 15.5.

15.6 Duality

To round out the chapter, we relate the discriminant and trace pairings to the dual and the different. For a detailed, general investigation of the dual in the context of other results for orders, see Faddeev [Fad65].

We continue with the hypothesis that R is a domain with $F = \operatorname{Frac} R$. Let B be an F-algebra with $n := \dim_F B < \infty$. As the trace pairing will play a significant role in what follows, we suppose throughout that B is *separable* (in particular, semisimple) as an F-algebra with reduced trace trd. Let I, J be R-lattices in B.

Definition 15.6.1. The **dual** of I (over R, with respect to trd) is

$$I^\sharp := \{\alpha \in B : \operatorname{trd}(\alpha I) \subseteq R\} = \{\alpha \in B : \operatorname{trd}(I\alpha) \subseteq R\}.$$

Some properties of the dual are evident.

Lemma 15.6.2.

(a) *If $I \subseteq J$ then $I^\sharp \supseteq J^\sharp$.*
(b) *For all $\beta \in B^\times$, we have $(\beta I)^\sharp = I^\sharp \beta^{-1}$.*
(c) *If $\mathfrak{p} \subseteq R$ is prime, then $(I_{(\mathfrak{p})})^\sharp = (I^\sharp)_{(\mathfrak{p})}$ and the same with the completion.*

Proof. For parts (a) and (b), see Exercise 15.15. The proof of part (c) is similarly straightforward. □

15.6.3. Suppose that I is free over R with basis $\alpha_1, \dots, \alpha_n$. Since the trace pairing on B is nondegenerate (Theorem 7.9.4), there exists a dual basis $\alpha_i^\sharp \in B$ to α_i under the reduced trace trd, so that $\operatorname{trd}(\alpha_i^\sharp \alpha_j) = 0, 1$ according as $i \neq j$ or $i = j$.

Then $I^\sharp$ is free over R with basis $\alpha_1^\sharp, \dots, \alpha_n^\sharp$: if $\beta = b_1\alpha_1^\sharp + \dots + b_n\alpha_n^\sharp$ with $b_1, \dots, b_n \in F$, then $\beta \in I^\sharp$ if and only if $\operatorname{trd}(\alpha_i\beta) = b_i \in R$ for all i.

Lemma 15.6.4. *$I^\sharp$ is an R-lattice in B.*

Proof. Let $\alpha_1, \dots, \alpha_n \in I$ be an F-basis for B, and let $J = \sum_i R\alpha_i \subseteq I$. Then there exists nonzero $r \in R$ such that $rI \subseteq J$, so $J \subseteq I \subseteq r^{-1}J$. Let $\alpha_1^\sharp, \dots, \alpha_n^\sharp \in B$ be the dual basis as in 15.6.3. It follows that $J^\sharp = \sum_i R\alpha_i^\sharp$ is an R-lattice, and consequently by Lemma 15.6.2(a)–(b) we have $rJ^\sharp \subseteq I^\sharp \subseteq J^\sharp$; since R is noetherian, $I^\sharp$ is an R-lattice. □

From now on, we suppose that R is a *Dedekind domain*; in particular, I is then projective as an R-module.

Lemma 15.6.5. *The natural inclusion $I \hookrightarrow (I^\sharp)^\sharp \subseteq B$ is an equality.*

Proof. If $\alpha \in I$ and $\beta \in I^\sharp$ then $\operatorname{trd}(\alpha\beta) \subseteq R$ and $\alpha \in (I^\sharp)^\sharp$. To show that the map is an equality, we argue locally, so we may suppose that I is free over R with basis α_i; then by applying 15.6.3 twice, $(I^\sharp)^\sharp$ has basis $(\alpha_i^\sharp)^\sharp = \alpha_i$, and equality holds. □

Proposition 15.6.6. *We have $O_{\mathsf{R}}(I) = O_{\mathsf{L}}(I^\sharp)$ and $O_{\mathsf{L}}(I) = O_{\mathsf{R}}(I^\sharp)$.*

Proof. First the inclusion ($\subseteq$). Let $\alpha \in O_{\mathsf{R}}(I)$; then $I\alpha \subseteq I$, so $I^\sharp I\alpha \subseteq I^\sharp I$ and

$$\operatorname{trd}(\alpha I^\sharp I) = \operatorname{trd}(I^\sharp I\alpha) \subseteq \operatorname{trd}(I^\sharp I) \subseteq R$$

hence $\alpha I^\sharp \subseteq I^\sharp$ and $\alpha \in O_{\mathsf{L}}(I^\sharp)$. Thus $O_{\mathsf{R}}(I) \subseteq O_{\mathsf{L}}(I^\sharp) \subseteq O_{\mathsf{R}}((I^\sharp)^\sharp) = O_{\mathsf{R}}(I)$ by Lemma 15.6.5, so equality holds. A similar argument works on the other side. □

The name *dual* is explained by the following lemma.

Proposition 15.6.7. *Let* $\operatorname{trd}(I) = \mathfrak{a} \subseteq F$*. Then the map*

$$\begin{aligned} I^\sharp &\xrightarrow{\sim} \operatorname{Hom}_R(I, \mathfrak{a}) \\ \beta &\mapsto (\alpha \mapsto \operatorname{trd}(\alpha\beta)) \end{aligned} \tag{15.6.8}$$

is an isomorphism of $O_{\mathsf{R}}(I), O_{\mathsf{L}}(I)$*-bimodules over* R*.*

Proof. For $\beta \in I$, let $\phi_\beta \colon I \to \mathfrak{a}$ be defined by $\phi_\beta(\alpha) = \operatorname{trd}(\alpha\beta)$ for $\alpha \in I$. The map $\beta \mapsto \phi_\beta \in \operatorname{Hom}_R(I, \mathfrak{a})$ from (15.6.8) is an R-module homomorphism. Moreover, it a map of $O_{\mathsf{R}}(I), O_{\mathsf{L}}(I)$-bimodules: if $\gamma \in O_{\mathsf{L}}(I)$ then $\gamma \in O_{\mathsf{R}}(I^\sharp)$ by Lemma 15.6.6, with induced map

$$\phi_{\beta\gamma}(\alpha) = \operatorname{trd}(\alpha\beta\gamma) = \operatorname{trd}(\gamma\alpha\beta) = \phi_\beta(\gamma\alpha) = (\gamma\phi_\beta)(\alpha) \tag{15.6.9}$$

and similarly on the other side.

Finally, we prove that the map (15.6.8) is also an isomorphism. Extending scalars to F, the trace pairing gives an isomorphism of F-vector spaces

$$\begin{aligned} \operatorname{Hom}_R(I, \mathfrak{a}) \otimes_R F \simeq \operatorname{Hom}_F(B, F) &\simeq B \\ \beta &\mapsto \phi_\beta \end{aligned}$$

because the pairing is nondegenerate (as B is separable). So immediately the map is injective; and it is surjective, because if $\phi \in \operatorname{Hom}_R(I, \mathfrak{a})$ then $\phi = \phi_\beta$ for some $\beta \in B$, but then $\phi(\alpha) = \operatorname{trd}(\alpha\beta) \in R$ for all $\alpha \in I$, so $\beta \in I^\sharp$ by definition. $\square$

Remark 15.6.10. The content of Proposition 15.6.7 is that although one can always construct the module dual, the trace pairing concretely realizes this module dual as a lattice. (And we speak of bimodules in the proposition because $\operatorname{Hom}_R(I, \mathfrak{a})$ does not come equipped with the structure of R-lattice in B.) This module duality, and the fact that I is projective over R, can be used to give another proof of Lemma 15.6.5.

The dual asks for elements that pair integrally under the trace. We might also ask for elements that multiply one lattice into another, as follows.

Definition 15.6.11. Let I, J be R-lattices. The **left colon lattice** of I with respect to J is the set

$$(I : J)_{\mathsf{L}} := \{\alpha \in B : \alpha J \subseteq I\}$$

and similarly the **right colon lattice** is

$$(I : J)_{\mathsf{R}} := \{\alpha \in B : J\alpha \subseteq I\}.$$

Note that $(I : I)_{\mathsf{L}} = O_{\mathsf{L}}(I)$ is the left order of I (and similarly on the right). The same proof as in Lemma 10.2.7 shows that $(I : J)_{\mathsf{L}}$ and $(I : J)_{\mathsf{R}}$ are R-lattices.

Lemma 15.6.12. *We have*

$$(IJ)^\sharp = (I^\sharp : J)_{\mathsf{R}} = (J^\sharp : I)_{\mathsf{L}}.$$

Proof. We have $\beta \in (IJ)^\sharp$ if and only if $\operatorname{trd}(\beta IJ) \subseteq R$ if and only if $\beta\alpha \in J^\sharp$ for all $\alpha \in I$ if and only if $\beta \in (J^\sharp : I)_{\mathsf{L}}$. A similar argument works on the other side, considering $\operatorname{trd}(IJ\beta)$ instead. □

Corollary 15.6.13. *We have* $\mathcal{O}_{\mathsf{L}}(I) = (II^\sharp)^\sharp$ *and* $\mathcal{O}_{\mathsf{R}}(I) = (I^\sharp I)^\sharp$.

Proof. Combining Lemmas 15.6.5 and 15.6.12,

$$\mathcal{O}_{\mathsf{L}}(I) = (I : I)_{\mathsf{L}} = ((I^\sharp)^\sharp : I)_{\mathsf{L}} = (II^\sharp)^\sharp$$

and similarly on the right. □

Definition 15.6.14. The **level** of I is the fractional ideal $\operatorname{lvl}(I) = \operatorname{nrd}(I^\sharp) \subseteq F$.

We now relate the above duality to the discriminant.

Definition 15.6.15. The **codifferent** of $\mathcal{O}$ is

$$\operatorname{codiff}(\mathcal{O}) := \mathcal{O}^\sharp.$$

Lemma 15.6.16. $\mathcal{O}_{\mathsf{L}}(\operatorname{codiff}(\mathcal{O})) = \mathcal{O}_{\mathsf{R}}(\operatorname{codiff}(\mathcal{O})) = \mathcal{O}$ *and* $\mathcal{O} \subseteq \operatorname{codiff}(\mathcal{O})$.

Proof. By Proposition 15.6.6, $\mathcal{O} = \mathcal{O}_{\mathsf{R}}(\mathcal{O}) = \mathcal{O}_{\mathsf{L}}(\operatorname{codiff}(\mathcal{O}))$ and similarly on the right. And $\mathcal{O} \subseteq \operatorname{codiff}(\mathcal{O})$ since $\operatorname{trd}(\mathcal{O}\mathcal{O}) = \operatorname{trd}(\mathcal{O}) \subseteq R$. □

The major role played by the codifferent is its relationship to the discriminant, as follows.

Lemma 15.6.17. $\operatorname{disc}(\mathcal{O}) = [\operatorname{codiff}(\mathcal{O}) : \mathcal{O}]_R$.

Proof. For a prime $\mathfrak{p} \subseteq R$ we have $\operatorname{disc}(\mathcal{O})_{(\mathfrak{p})} = \operatorname{disc}(\mathcal{O}_{(\mathfrak{p})})$ and $[\mathcal{O}^\sharp_{(\mathfrak{p})} : \mathcal{O}_{(\mathfrak{p})}]_{R_{(\mathfrak{p})}} = ([\mathcal{O}^\sharp : \mathcal{O}]_R)_{(\mathfrak{p})}$, and so to establish the equality we may argue locally. Since $\mathcal{O}_{(\mathfrak{p})}$ is free over $R_{(\mathfrak{p})}$, we reduce to the case where $\mathcal{O}$ is free over R, say $\mathcal{O} = \sum_i R\alpha_i$. Then $\mathcal{O}^\sharp = \sum_i R\alpha_i^\sharp$ with $\alpha_1^\sharp, \dots, \alpha_n^\sharp \in B$ the dual basis, as in 15.6.3.

The ideal $\operatorname{disc}(\mathcal{O})$ is principal, generated by $d(\alpha_1, \dots, \alpha_n) = \det(\operatorname{trd}(\alpha_i\alpha_j))_{i,j}$; at the same time, the R-index $[\mathcal{O}^\sharp : \mathcal{O}]_R$ is generated by $\det(\delta)$ where δ is the change of basis from $\alpha_i^\sharp$ to α_i. But δ is precisely the matrix $(\operatorname{trd}(\alpha_i\alpha_j))_{i,j}$ (Exercise 15.14), and the result follows. □

Remark 15.6.18. In certain circumstances, it is preferable to work with an integral ideal measuring the discriminant, so instead of the codifferent instead a *different*: we will want to take a kind of inverse. We study this in the next chapter: see section 16.8.

Exercises

Unless otherwise specified, let R be a noetherian domain with field of fractions F.

1. Let $\operatorname{char} F = 2$ and let $\left[\dfrac{a,b}{F}\right)$ be a quaternion algebra over F with $a, b \in R$ and $b \neq 0$. Show that $\mathcal{O} = R + Ri + Rj + Rij$ is an R-order in B and compute the (reduced) discriminant of $\mathcal{O}$.
2. Let $B = \mathrm{M}_n(F)$ and $\mathcal{O} = \mathrm{M}_n(R)$ with $n \geq 1$. Show that $\operatorname{disc}(\mathcal{O}) = R$. *[Hint: Compute directly on a basis $\{e_{ij}\}_{i,j}$ of matrix units, which satisfy $e_{ij}e_{i'j'} = e_{ij'}$ if $j = i'$, otherwise zero.]*
3. Suppose R is a global ring, so F is a global field; let B be a quaternion algebra over F and let $\mathcal{O} \subseteq B$ be an R-order. Prove that for all primes $\mathfrak{p} \subseteq R$, we have $\mathcal{O}_{\mathfrak{p}} \simeq \mathrm{M}_n(R_{\mathfrak{p}})$ if and only if $\mathfrak{p} \nmid \operatorname{disc} \mathcal{O}$.
4. Let $B := (K, b \mid F)$ be a quaternion algebra over a field F with $b \in F^\times$. Let $S \subseteq K$ be an R-order with $\mathfrak{d} := \operatorname{disc}(S)$; let $\mathfrak{b} \subseteq K$ be a fractional S-ideal (which can be but need not be invertible), and finally let $\mathcal{O} := S \oplus \mathfrak{b}j$.

   (a) Show that $\mathcal{O}$ is an R-order if and only if $\operatorname{Nm}_{K|F} \mathfrak{b} \subseteq b^{-1}R$.
   (b) Compute that $\operatorname{discrd} \mathcal{O} = \mathfrak{d}(\operatorname{Nm}_{K|F} \mathfrak{b})b$.

5. In this exercise, we consider a construction of maximal orders as crossed products in the simplest case over $\mathbb{Q}$, continuing Exercise 14.9. Let $B := \left(\dfrac{q^\diamond, b}{\mathbb{Q}}\right)$ be a quaternion algebra of discriminant D, where $b \in \mathbb{Z}$ is squarefree with $D \mid b$ and q is an odd prime with $q^\diamond = \pm q \equiv 1 \pmod 4$, the minus sign if B is indefinite.

   Let $K := \mathbb{Q}(\sqrt{q^\diamond})$ be the quadratic field of discriminant $q^\diamond$. Let $S \subseteq K$ be the ring of integers of K, so $\operatorname{disc} S = q^\diamond$.

   (a) Show that for all odd primes $p \mid (b/D)$, we have $\left(\dfrac{q^\diamond}{p}\right) = 1$. Conclude there exists an ideal $\mathfrak{b} \subseteq S$ such that $\operatorname{Nm} \mathfrak{b} = b/D$.
   (b) Let $\mathfrak{q} \subseteq S$ be the unique prime above q, and let
   $$\mathcal{O} := S \oplus (\mathfrak{q}\mathfrak{b})^{-1}j$$
   Show that $\mathcal{O}$ is a maximal order in B.
   (c) Let $c \in \mathbb{Z}$ satisfy $c^2 \equiv q^\diamond \pmod{4b/D}$. Show that the order $\mathcal{O}$ in (b) can be written
   $$\mathcal{O} = \mathbb{Z} \oplus \mathbb{Z}\frac{1+i}{2} \oplus \mathbb{Z}j \oplus \mathbb{Z}\frac{D(c+i)j}{2bq}.$$

6. Let B be a separable F-algebra with $\dim_F B = n$. Show that $\alpha_1, \dots, \alpha_n \in B$ are linearly independent over F if and only if $d(\alpha_1, \dots, \alpha_n) \neq 0$.
7. Let $\mathcal{O}$ be an R-order. Show that $\operatorname{disc}(\mathcal{O})$ is generated by
$$\{d(1, \alpha_1, \dots, \alpha_{n-1}) : \alpha_1, \dots, \alpha_{n-1} \in \mathcal{O}\}.$$

8. Let I be an R-lattice in B over F, let K be a finite extension field of F, and let S be a domain containing R with field of fractions K. Show that
$$\operatorname{disc}(I \otimes_R S) = \operatorname{disc}(I) \otimes_R S = \operatorname{disc}(I)S.$$

9. Let $\operatorname{char} F \neq 2$ and let B be a quaternion algebra over F. Let $\alpha, \beta \in B$ be such that $F(\alpha) \cap F(\beta) = F$. Recall the discriminant form Δ (Exercise 4.3), and let
$$s := \operatorname{trd}(\alpha\beta) - \frac{\operatorname{trd}(\alpha)\operatorname{trd}(\beta)}{2}.$$
Show that
$$d(1, \alpha, \beta, \alpha\beta) = -(s^2 - 4\Delta(\alpha)\Delta(\beta))^2 = -\Delta(\alpha\beta)^2.$$
[Hint: reduce to the case where $\operatorname{trd}(\alpha) = \operatorname{trd}(\beta) = 0$, *noting the invariance of* s.*]*

10. Let B be a quaternion algebra over F. Define $m : B \times B \times B \to F$ by $m(\alpha_1, \alpha_2, \alpha_3) := \operatorname{trd}([\alpha_1, \alpha_2]\overline{\alpha_3})$ for $\alpha_i \in B$. If $\beta_i = M\alpha_i$ for some $M \in \mathrm{M}_3(F)$, show that
$$m(\beta_1, \beta_2, \beta_3) = \det(M) m(\alpha_1, \alpha_2, \alpha_3).$$

11. Let B be a quaternion algebra over F. Give another proof that
$$m(\alpha_1, \alpha_2, \alpha_3)^2 = d(1, \alpha_1, \alpha_2, \alpha_3)$$
(cf. Brzezinski [Brz82, Lemma 1.1(a)]) for all $\alpha_i \in B$ as follows:

   (a) Suppose $B = \mathrm{M}_2(F)$. Show that the matrix units
$$e_{12} = \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}, \quad e_{21} = \begin{pmatrix} 0 & 0 \\ 1 & 0 \end{pmatrix}, \quad e_{22} = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}$$
   span B/F, and $m(e_{12}, e_{21}, e_{22})^2 = d(1, e_{12}, e_{21}, e_{22})$. Conclude using Exercise 15.10.

   (b) Reduce to (a) in general by taking a splitting field for B.

12. Suppose $R = R_{(S)}$ is a global ring with $2 \in R^\times$. Let $K \supset F$ be a quadratic field extension and $S \subseteq K$ an R-order. Let $\operatorname{Ram}(K)$ be the set of places of F that are ramified in K. Show that S is maximal if and only if its discriminant is equal to
$$\operatorname{disc}_R(S) = \prod_{\mathfrak{p} \in \operatorname{Ram}(K) \setminus S} \mathfrak{p} \subseteq R$$
in analogy with Theorem 15.5.5.

13. Let B be a finite-dimensional F-algebra with a standard involution. Compare
$$\det(\operatorname{trd}(\alpha_i\alpha_j))_{i,j} \quad \text{with} \quad \det(\operatorname{trd}(\alpha_i\overline{\alpha_j}))_{i,j}$$
for $\alpha_i \in B$, and show that defining the discriminant of an order $O \subseteq B$ with respect to either pairing gives the same result.

▹ 14. Let B be a semisimple F-algebra with $\dim_F B = n$, let I be an R-lattice that is free over R with basis $\alpha_1, \dots, \alpha_n$, and let $\alpha_1^\sharp, \dots, \alpha_n^\sharp \in B$ be the dual basis, so $\operatorname{trd}(\alpha_i^\sharp \alpha_j) = 1, 0$ according as $i = j$ or not. Show that the change of basis matrix from $\{\alpha_i^\sharp\}_i$ to $\{\alpha_i\}_i$ is given by $(\operatorname{trd}(\alpha_i\alpha_j))_{i,j}$.

15. Let $I \subseteq B$ be an R-lattice in a separable algebra B.

   (a) If $J \subseteq B$ is an R-lattice with $I \subseteq J$, show that $I^\sharp \supseteq J^\sharp$.
   (b) Show that for all $\beta \in B^\times$, we have $(\beta I)^\sharp = I^\sharp \beta^{-1}$

16. Let R be a DVR with maximal ideal $\mathfrak{p} = \pi R$, and let $O := \begin{pmatrix} R & R \\ \mathfrak{p}^e & R \end{pmatrix}$ for $e \geq 0$. Compute the codifferent $\operatorname{codiff}(O)$: in particular, show that $\operatorname{codiff}(O)$ is a principal two-sided O-ideal, and find a generator. Verify Lemma 15.6.17.

17. Let R be a noetherian domain with $F = \operatorname{Frac} R$. Let B be a central simple algebra over F. Let $O \subseteq B$ be an R-order. We say O is **Azumaya** if O is R-**simple**, which is to say every two-sided ideal $I \subseteq O$ is of the form $\mathfrak{a}O = O\mathfrak{a}$ with $\mathfrak{a} = I \cap R \subseteq R$.

   (a) Show that O is Azumaya if and only if every R-algebra homomorphism $O \to A$ is either the zero map or injective.
   (b) Show that O is Azumaya if and only if $O/\mathfrak{m}O$ is a central simple algebra over the field $R/\mathfrak{m}$ for all maximal ideals $\mathfrak{m}$ of R.
   (c) Suppose that B is a quaternion algebra. Show that the quaternion order O is Azumaya if and only if $\operatorname{disc} O = R$. Conclude that the only Azumaya quaternion algebra over the valuation ring R of a local field is $M_2(R)$, and that the only Azumaya quaternion algebra over $\mathbb{Z}$ is $M_2(\mathbb{Z})$.

   [See Auslander and Goldman [AG60] or Milne [Milne80, §IV.1].]

18. Let G be a finite group of order $n = \#G$ and let R be a domain with $F = \operatorname{Frac} R$. Suppose that $\operatorname{char} F \nmid n$. Then $B := F[G]$ is a separable F-algebra by Exercise 17.6.

   (a) Consider the algebra trace $\operatorname{Tr}_{B|F}$ and its associated bilinear form. Show that in the basis of $F[G]$ given by the elements of G that the trace pairing is the scalar matrix n.
   (b) Now write $B \simeq B_1 \times \cdots \times B_r$ as a product of simple F-algebras. Let K_i be the center of B_i, and let $\dim_{K_i} B_i = n_i^2$. Show that $\operatorname{Tr}|_{B_i} = n_i \operatorname{trd}$. Let $O = R[G]$, and suppose that $O \simeq O_1 \times \cdots \times O_r$. Show that

$$\operatorname{codiff}(O) = n_1^{-1} O_1 \times \cdots \times n_r^{-1} O_r.$$

This chapter is licensed under the terms of the Creative Commons Attribution-NonCommercial 4.0 International License (http://creativecommons.org/licenses/by-nc/4.0/), which permits any noncommercial use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license and indicate if changes were made.

The images or other third party material in this chapter are included in the chapter's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the chapter's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.

Chapter 16

Quaternion ideals and invertibility

Much like a space can be understood by studying functions on that space, often the first task to understand a ring A is to understand the ideals of A and modules over A (in other words, to pursue "linear algebra" over A). The ideals of a ring that are easiest to work with are the principal ideals—but not all ideals are principal, and various algebraic structures are built to understand the difference between these two. In this chapter, we consider these questions for the case where A is a quaternion order.

16.1 ▹ Quaternion ideals

To get warmed up for the noncommutative situation, we consider ideals of quadratic rings. An integer $d \in \mathbb{Z}$ is a **discriminant** if $d \equiv 0, 1 \pmod 4$. Let S be the quadratic order of nonsquare discriminant $d \in \mathbb{Z}$, namely,

$$S = S(d) := \mathbb{Z} \oplus \mathbb{Z}[(d + \sqrt{d})/2] \subset K = \mathbb{Q}(\sqrt{d}).$$

The set of ideals of S has a natural multiplicative structure with identity element S (giving it the structure of a commutative monoid), but we lack inverses and we would surely feel more comfortable with a group structure. So we consider nonzero S-lattices $\mathfrak{a} \subset K$, and call them **fractional ideals** of S; equivalently, they are the S-submodules $d^{-1}\mathfrak{a} \subset K$ with $\mathfrak{a} \subseteq S$ a nonzero ideal and $d \in \mathbb{Z}_{>0}$, hence the name *fractional ideal* (viz. 9.2.4). To get a group structure, we must restrict our attention to the **invertible** fractional ideals $\mathfrak{a} \subset K$, i.e., those such that there exists a fractional ideal $\mathfrak{b}$ with $\mathfrak{a}\mathfrak{b} = S$. The simplest kind of invertible fractional ideals are the principal ones $\mathfrak{a} = aS$ for $a \in K^\times$, with inverse $\mathfrak{a}^{-1} = a^{-1}S$. If a fractional ideal $\mathfrak{a}$ has an inverse then this inverse is unique, given by

$$\mathfrak{a}^{-1} = \{x \in K : x\mathfrak{a} \subseteq S\};$$

and for a fractional ideal $\mathfrak{a}$, we always have $\mathfrak{a}\mathfrak{a}^{-1} \subseteq S$ (but equality may not hold). If $S = \mathbb{Z}_K$ is the ring of integers (the maximal order) of K, then all nonzero fractional ideals of S are invertible—in fact, this property characterizes Dedekind domains, in that a noetherian commutative ring is a Dedekind domain if and only if every nonzero (prime) ideal is invertible. (See also the summary in section 9.2.)

© The Author(s) 2021

J. Voight, *Quaternion Algebras*, Graduate Texts in Mathematics 288,
https://doi.org/10.1007/978-3-030-56694-4_16

A fractional ideal $\mathfrak{a}$ of S is invertible if and only if $\mathfrak{a}$ is **locally principal**, i.e., $\mathfrak{a} \otimes_{\mathbb{Z}} \mathbb{Z}_{(p)} = \mathfrak{a}_{(p)} = a_{(p)}\mathbb{Z}_{(p)}$ is a principal fractional ideal of the localization $S_{(p)}$ for all primes p. Every locally principal ideal is invertible, and the extent to which the converse holds is something that arises in an important way more generally in algebraic geometry. In the language of commutative algebra, a locally principal S-module is equivalently a projective S-module of rank 1.

Suppose that S is not maximal; then $S(d) = \mathbb{Z} + f\mathbb{Z}_K$ for a unique integer $f \in \mathbb{Z}_{>1}$, the **conductor** of S; accordingly, $d = d_K f^2$, where $d_K \in \mathbb{Z}$ is the discriminant of $\mathbb{Z}_K$ (a **fundamental discriminant**). In this case, there is always an ideal of S that is not invertible. Specifically, consider the ideal

$$\mathfrak{f} = f\mathbb{Z} + \sqrt{d}\mathbb{Z} \subseteq S. \tag{16.1.1}$$

Then $\mathfrak{f}$ is a free $\mathbb{Z}$-module of rank 2 and

$$\mathfrak{f}^2 = (f\mathbb{Z} + \mathbb{Z}\sqrt{d})^2 = f^2\mathbb{Z} + f\sqrt{d}\mathbb{Z} = f\mathfrak{f}$$

so if $\mathfrak{f}$ were invertible, then cancelling we would obtain $\mathfrak{f} = fS$, a contradiction. The source of this example is that $\mathfrak{f} = fS(d_K)$ since $S(d_K) = \mathbb{Z} + \sqrt{d_K}\mathbb{Z}$, so really this fractional ideal belongs to the maximal order $S(d_K)$, not to S. For more on the notion of invertibility for quadratic orders, see Cox [Cox89, §7], with further connections to quadratic forms and class numbers.

We now turn to the quaternionic generalization, where noncommutativity presents some complications. Let B be a quaternion algebra over $\mathbb{Q}$ and let $O \subset B$ be an order. To study ideals of O we must distinguish between left or right ideals, and the product of two (say) right O-ideals need not be again a right O-ideal! To address this, for lattices $I, J \subset B$, we say that I is **compatible** with J if the right order of I is equal to the left order of J, so that what comes between I and J in the product $I \cdot J$ "matches up".

A lattice $I \subset B$ is **right invertible** if there exists a lattice $I' \subset B$ such that

$$II' = O_{\mathsf{L}}(I)$$

with a compatible product, and we call I' a **right inverse**. We similarly define notions on the left, and we say $I \subset B$ is **invertible** if there is a two-sided inverse $I' \subset B$, so

$$II' = O_{\mathsf{L}}(I) = O_{\mathsf{R}}(I') \text{ and } I'I = O_{\mathsf{L}}(I') = O_{\mathsf{R}}(I)$$

with both of these products are compatible. If a lattice I has a two-sided inverse, then this inverse is uniquely given by

$$I^{-1} := \{\alpha \in B : I\alpha I \subseteq I\}$$

(defined so as to simultaneously take care of both left and right): we always have that $II^{-1} \subseteq O_{\mathsf{L}}(I)$, but *equality* is needed for right invertibility, and the same on the left.

Let $O \subseteq B$ be an order. A **left fractional O-ideal** is a lattice $I \subseteq B$ such that $O \subseteq O_{\mathsf{L}}(I)$; we similarly define on the right. For a maximal order, all lattices are invertible (Proposition 16.6.15(b)).

Proposition 16.1.2. *Let $O \subseteq B$ be a maximal order. Then a left or right fractional O-ideal is invertible.*

The simplest kind of invertible lattices are the **principal** lattices

$$I = O_{\mathsf{L}}(I)\alpha = \alpha O_{\mathsf{R}}(I)$$

with $\alpha \in B^\times$: its inverse is $I^{-1} = \alpha^{-1}O_{\mathsf{L}}(I) = O_{\mathsf{R}}(I)\alpha^{-1}$.

The major task of this chapter will be to interrelate these notions in the quaternionic context. Let

$$\operatorname{nrd}(I) := \gcd(\{\operatorname{nrd}(\alpha) : \alpha \in I\}),$$

i.e., $\operatorname{nrd}(I)$ is a positive generator of the (finitely generated) subgroup of $\mathbb{Q}$ generated by $\operatorname{nrd}(\alpha)$ for $\alpha \in I$. The main result over $\mathbb{Q}$ is the following theorem (Main Theorem 16.7.7).

Main Theorem 16.1.3. *Let B be a quaternion algebra over $\mathbb{Q}$ and let $I \subset B$ be an integral lattice. Then the following are equivalent:*

(i) *I is locally principal, i.e., $I_{(p)} = I \otimes_{\mathbb{Z}} \mathbb{Z}_{(p)}$ is principal for all primes p;*
(ii) *I is invertible;*
(iii) *I is right invertible;*
(iii′) *I is left invertible;*
(iv) $\operatorname{nrd}(I)^2 = [O_{\mathsf{R}}(I) : I]$*; and*
(iv′) $\operatorname{nrd}(I)^2 = [O_{\mathsf{L}}(I) : I]$.

Accordingly, for I integral, we may define the **right absolute norm** of I by

$$\mathsf{N}(I) := \#(O_{\mathsf{R}}(I)/I) = [O_{\mathsf{R}}(I) : I] \in \mathbb{Z}_{\geq 1}$$

and similarly on the left; by Main Theorem 16.1.3 (iv) $\Leftrightarrow$ (iv′), when I is locally principal, the left and right absolute norms coincide (called then just **absolute norm**) and are related to the reduced norm by $\mathsf{N}(I) = \operatorname{nrd}(I)^2$.

16.2 Locally principal, compatible lattices

The simplest lattices to understand are those that are principal; but as we saw in section 9.4, lattices over Dedekind domains are inherently local in nature. We are led to consider the more general class of *locally principal* lattices. We work first with lattices, and later we will keep track of their left and right orders.

Throughout this chapter, let R be a Dedekind domain with field of fractions F, let B be a finite-dimensional algebra over F, and let $I \subseteq B$ be an R-lattice.

Definition 16.2.1. I is **principal** if there exists $\alpha \in B$ such that

$$I = O_{\mathsf{L}}(I)\alpha = \alpha O_{\mathsf{R}}(I);$$

we say that I is **generated** by α.

16.2.2. If I is generated by $\alpha \in B$, then since I is a lattice (Definition 9.3.1) we have $IF = B\alpha = B$, so $\alpha \in B^\times$.

16.2.3. If $I = O_{\mathsf{L}}(I)\alpha$, then $O_{\mathsf{R}}(I) = \alpha^{-1}O_{\mathsf{L}}(I)\alpha$ by Exercise 16.2, so

$$I = \alpha(\alpha^{-1}O_{\mathsf{L}}(I)\alpha) = \alpha O_{\mathsf{R}}(I).$$

Therefore it is sufficient to check for a one-sided generator (and if we defined the obvious notions of **left principal** or **right principal**, these would be equivalent to the notion of principal).

The notion of principality naturally extends locally.

Definition 16.2.4. An R-lattice I is **locally principal** if $I_{(\mathfrak{p})} = I \otimes_R R_{(\mathfrak{p})}$ is a principal $R_{(\mathfrak{p})}$-lattice for all primes $\mathfrak{p}$ of R.

Now let I, J be R-lattices in B. We define the product IJ to be the R-submodule of B generated by the set

$$\{\alpha\beta : \alpha \in I,\ \beta \in J\}.$$

The product IJ is an R-lattice: it is finitely generated as this is true of I, J individually, and there exists a nonzero $r \in I$ (Exercise 9.2) so $rJ \subset IJ$ and thus

$$B = FJ = F(rJ) \subseteq FIJ$$

so equality holds.

When multiplication of two lattices matches up their respective left and right orders, we give it a name.

Definition 16.2.5. We say that I is **compatible** with J if $O_{\mathsf{R}}(I) = O_{\mathsf{L}}(J)$.

We will also sometimes just say that the product IJ is **compatible** to mean that I is compatible with J. The relation "is compatible with" is in general neither symmetric nor transitive.

16.2.6. I has the structure of a right $O_{\mathsf{R}}(I)$-module and J the structure of a left $O_{\mathsf{L}}(J)$-module. When $O_{\mathsf{R}}(I) = O_{\mathsf{L}}(J) = O$, that is, when I is compatible with J, it makes sense to consider the tensor product $I \otimes_O J$ as an R-module. The multiplication map $B \otimes_B B \xrightarrow{\sim} B$ defined by $\alpha \otimes \beta \mapsto \alpha\beta$ restricts to give an isomorphism $I \otimes_O J \xrightarrow{\sim} IJ$ as R-lattices. In this way, multiplication of compatible lattices can be thought of as a special case of the tensor product of modules.

We conclude this section with several other basic properties of lattices.

Definition 16.2.7. An R-lattice I is **integral** if $I^2 \subseteq I$.

In Definition 16.2.7, the product need not be compatible.

Lemma 16.2.8. *Let I be an R-lattice. Then the following are equivalent:*

(i) *I is integral;*

(ii) *For all* $\alpha, \beta \in I$, *we have* $\alpha\beta \in I$;
(iii) $I \subseteq \mathcal{O}_{\mathsf{L}}(I)$, *so* I *is a left ideal of* $\mathcal{O}_{\mathsf{L}}(I)$ *in the usual sense;*
(iii′) $I \subseteq \mathcal{O}_{\mathsf{R}}(I)$; *and*
(iv) $I \subseteq \mathcal{O}_{\mathsf{L}}(I) \cap \mathcal{O}_{\mathsf{R}}(I)$.

If I *is integral, then every element of* I *is integral over* R.

Proof. The equivalence (i) ⇔ (ii) follows immediately. For (i) ⇔ (iii), we have $II \subseteq I$ if and only if $I \subseteq \mathcal{O}_{\mathsf{L}}(I)$ by definition of $\mathcal{O}_{\mathsf{L}}(I)$, and the same argument gives (i) ⇔ (iii′), and this then gives (i) ⇔ (iv). The final statement follows from Lemma 10.3.2. □

In light of Lemma 16.2.8, we need not define notions of *left integral* or *right integral*.

For an R-lattice I, there exists nonzero $d \in R$ such that dI is integral, so every R-lattice $I = (dI)/d$ is *fractional* in the sense that it is obtained from an integral lattice with denominator.

Definition 16.2.9. Let $\mathcal{O} \subseteq B$ be an R-order. A **left fractional $\mathcal{O}$-ideal** is a lattice $I \subseteq B$ such that $\mathcal{O} \subseteq \mathcal{O}_{\mathsf{L}}(I)$; similarly on the right.

If $\mathcal{O}, \mathcal{O}' \subseteq B$ are R-orders, then a **fractional $\mathcal{O}, \mathcal{O}'$-ideal** is a lattice I that is a left fractional $\mathcal{O}$-ideal and a right fractional $\mathcal{O}'$-ideal.

Remark 16.2.10. A left ideal $I \subseteq \mathcal{O}$ in the usual sense is an integral left $\mathcal{O}$-ideal in the sense of Definition 16.2.9 if and only if $IF = B$, i.e., I is a (full) R-lattice. (Same for right and two-sided ideals.) If I is nonzero and B is a division algebra, then automatically I is full and the two notions coincide.

Indeed, suppose $I \subseteq \mathcal{O}$ is a left ideal of $\mathcal{O}$ (in the usual sense). Then $\mathcal{O} \subseteq \mathcal{O}_{\mathsf{L}}(I)$ so in particular I has the structure of an R-module, and since $\mathcal{O}$ is finitely generated as an R-module and R is noetherian, it follows that I is finitely generated. Consequently, a left ideal $I \subseteq \mathcal{O}$ is a left fractional $\mathcal{O}$-ideal if and only if $IF = B$.

Definition 16.2.11. Let I be a left fractional $\mathcal{O}$-ideal. We say that I is **sated** (as a left fractional $\mathcal{O}$-ideal) if $\mathcal{O} = \mathcal{O}_{\mathsf{L}}(I)$. We make a similar definition on the right and for two-sided ideals.

Example 16.2.12. By Lemma 15.6.16, $\operatorname{codiff}(\mathcal{O})$ is a two-sided sated $\mathcal{O}$-ideal.

Remark 16.2.13. Our notion of *sated* is sometimes called *proper*: we do not use this already overloaded term, as it conflicts with the notion of a proper subset.

16.3 Reduced norms

Next, we extend the reduced norm to lattices; see also Reiner [Rei2003, §24]. To this end, in this section we suppose that B is semisimple.

Definition 16.3.1. The **reduced norm** $\operatorname{nrd}(I)$ of I is the R-submodule of F generated by the set $\{\operatorname{nrd}(\alpha) : \alpha \in I\}$.

Lemma 16.3.2. *The reduced norm* $\operatorname{nrd}(I)$ *is a fractional ideal of* F*: i.e., it is finitely generated as an* R*-module.*

Proof. We first give a proof when B has a standard involution, and nrd is a quadratic form. Since I is an R-lattice we have $IF = B$; since $\operatorname{nrd}(B) \neq \{0\}$, we have $\operatorname{nrd}(I) \neq \{0\}$. And I is generated by finitely many α_i as an R-module; the R-module $\operatorname{nrd}(I)$ is then generated by the values $a_{ii} = \operatorname{nrd}(\alpha_i)$ and $a_{ij} = \operatorname{nrd}(\alpha_i + \alpha_j) - \operatorname{nrd}(\alpha_i) - \operatorname{nrd}(\alpha_j)$, since then

$$\operatorname{nrd}\left(\textstyle\sum_i c_i\alpha_i\right) = \textstyle\sum_{i,j} a_{ij}c_ic_j \in \textstyle\sum_{i,j} Ra_{ij}$$

for all $c_i \in R$.

Now for the general case. Replacing I by rI with $r \in R$ nonzero, we may suppose that I is integral, and hence $\operatorname{nrd}(I) \subseteq R$. Since I is a lattice, there exists $r \in I \cap R$ with $r \neq 0$. For all $\mathfrak{p}$ such that $\operatorname{ord}_{\mathfrak{p}}(r) = 0$, we have $1 \in I_{(\mathfrak{p})}$ so $\operatorname{nrd}(I_{(\mathfrak{p})}) = R_{(\mathfrak{p})}$. For each of the finitely many primes $\mathfrak{p}$ that remain, we choose an element $\alpha \in I$ such that $\operatorname{ord}_{\mathfrak{p}}(\operatorname{nrd}(\alpha))$ is minimal; then $\operatorname{nrd}(\alpha)$ generates $\operatorname{nrd}(I_{(\mathfrak{p})})$, and by the local-global dictionary, these finitely many elements generate $\operatorname{nrd}(I)$. ◻

16.3.3. For a prime $\mathfrak{p}$ of R we have $\operatorname{nrd}(I)_{(\mathfrak{p})} = \operatorname{nrd}(I_{(\mathfrak{p})})$, so by the local-global property of lattices (Lemma 9.4.6),

$$\operatorname{nrd}(I) = \bigcap_{\mathfrak{p}} \operatorname{nrd}(I)_{(\mathfrak{p})} = \bigcap_{\mathfrak{p}} \operatorname{nrd}(I_{(\mathfrak{p})}). \tag{16.3.4}$$

16.3.5. If I is a principal R-lattice generated by $\alpha \in I$ then $\operatorname{nrd}(I) = \operatorname{nrd}(\alpha)R$; more generally, if I is an R-lattice and $\alpha \in B^\times$ then $\operatorname{nrd}(\alpha I) = \operatorname{nrd}(\alpha)\operatorname{nrd}(I)$ (Exercise 16.4).

Now suppose that I, J are lattices. Then $\operatorname{nrd}(IJ) \supseteq \operatorname{nrd}(I)\operatorname{nrd}(J)$. However, we need not have equality, as the following example indicates.

Example 16.3.6. It is not always true that $\operatorname{nrd}(IJ) = \operatorname{nrd}(I)\operatorname{nrd}(J)$. For example, if $a \in R$ is neither zero nor a unit, then $I = \begin{pmatrix} aR & R \\ aR & R \end{pmatrix}$ and $J = \begin{pmatrix} aR & aR \\ R & R \end{pmatrix}$ are R-lattices in $\mathrm{M}_2(F)$ with $\operatorname{nrd}(I) = \operatorname{nrd}(J) = aR$ but $IJ = \mathrm{M}_2(R)$ and so $\operatorname{nrd}(IJ) = R$.

We have $O_{\mathsf{R}}(J) = \mathrm{M}_2(R) = O_{\mathsf{L}}(I)$, so J is compatible with I, and $\operatorname{nrd}(JI) = a^2R = \operatorname{nrd}(J)\operatorname{nrd}(I)$; but

$$O_{\mathsf{R}}(I) = \begin{pmatrix} R & a^{-1}R \\ aR & R \end{pmatrix} \quad \text{and} \quad O_{\mathsf{L}}(J) = \begin{pmatrix} R & aR \\ a^{-1}R & R \end{pmatrix},$$

so I is not compatible with J.

The issue present in Example 16.3.6 is that the product is not as well-behaved for noncommutative rings as for commutative rings; we need the elements coming between I and J to match up.

Lemma 16.3.7. *Suppose that* I *is compatible with* J *and that either* I *or* J *is locally principal. Then* $\operatorname{nrd}(IJ) = \operatorname{nrd}(I)\operatorname{nrd}(J)$.

Proof. By the local-global property for norms (16.3.4) and since localization commutes with multiplication, i.e.,

$$(\mathfrak{a}\mathfrak{b})_{(\mathfrak{p})} = \mathfrak{a}_{(\mathfrak{p})}\mathfrak{b}_{(\mathfrak{p})} \text{ for all (finitely generated) } R\text{-modules } \mathfrak{a}, \mathfrak{b} \subseteq F,$$

we may localize and suppose that either I or J is principal. Suppose I is (right) principal. Then $I = \alpha O$ for some $\alpha \in B$ where $O = O_{\mathsf{R}}(I) = O_{\mathsf{L}}(J)$. Then

$$IJ = (\alpha O)J = \alpha(OJ) = \alpha J$$

and so $\operatorname{nrd}(IJ) = \operatorname{nrd}(\alpha)\operatorname{nrd}(J) = \operatorname{nrd}(I)\operatorname{nrd}(J)$ by 16.3.5. The case where J is principal follows in the same way. □

Principal lattices are characterized by reduced norms, as follows.

Lemma 16.3.8. *Let I be locally principal and let $\alpha \in I$. Then α generates I if and only if* $\operatorname{nrd}(\alpha)R = \operatorname{nrd}(I)$.

Proof. If $I = \alpha O$ then $\operatorname{nrd}(I) = \operatorname{nrd}(\alpha)R$ by Lemma 16.3.7.

For the converse, let $O = O_{\mathsf{R}}(I)$. We want to show that $I = \alpha O$, and we know that $I \supseteq \alpha O$. To prove that equality holds, it suffices to show this locally, so we may suppose that $I = \beta O$. Then $\alpha = \beta\mu$ with $\mu \in O$, and $\operatorname{nrd}(\alpha) = \operatorname{nrd}(\beta\mu) = \operatorname{nrd}(\beta)\operatorname{nrd}(\mu)$. By hypothesis, $\operatorname{nrd}(\mu) \in R^\times$, and thus $\mu \in O^\times$, so $\beta O = \alpha O$. □

16.4 Algebra and absolute norm

The reduced norm of an ideal is related to its algebra norm, as follows. We continue to suppose that B is semisimple, so the definitions of left and right norm coincide.

Definition 16.4.1. The **(algebra) norm** $\operatorname{Nm}_{B|F}(I)$ of I is the R-submodule of F generated by the set $\{\operatorname{Nm}_{B|F}(\alpha) : \alpha \in I\}$.

Remark 16.4.2. The definition of algebra norm by necessity depends on the choice of domain R; indeed, I is an R-lattice.

Proposition 16.4.3. *The following are equivalent:*

(i) *I is locally principal;*

(ii) $\operatorname{Nm}_{B|F}(I) = [O_{\mathsf{L}}(I) : I]_R$*; and*

(iii) $\operatorname{Nm}_{B|F}(I) = [O_{\mathsf{R}}(I) : I]_R$.

If B is simple with $\dim_F B = n^2$*, then these are further equivalent to*

(iv) $\operatorname{nrd}(I)^n = [O_{\mathsf{L}}(I) : I]_R$.

(v) $\operatorname{nrd}(I)^n = [O_{\mathsf{R}}(I) : I]_R$.

Proof. Let $\mathcal{O} = \mathcal{O}_{\mathsf{L}}(I)$. Let $\alpha \in I$. Right multiplication by α gives an R-module isomorphism $\mathcal{O} \xrightarrow{\sim} \mathcal{O}\alpha$ (change of basis between two free R-modules), so by Lemma 9.6.4 we have $[\mathcal{O} : \mathcal{O}\alpha]_R = \det(\alpha)R = \mathrm{Nm}_{B|F}(\alpha)R$, thinking of $\alpha \in \mathrm{End}_F(B)$.

We now prove (i) $\Leftrightarrow$ (ii). We may suppose R is local, so R is a DVR. For all $\alpha \in I$, the following equality holds:

$$[\mathcal{O} : I]_R[I : \mathcal{O}\alpha]_R = [\mathcal{O} : \mathcal{O}\alpha]_R = \mathrm{Nm}_{B|F}(\alpha)R. \tag{16.4.4}$$

To show (i) $\Rightarrow$ (ii), if $I = \mathcal{O}\alpha$ then $\mathrm{Nm}_{B|F}(I) = \mathrm{Nm}_{B|F}(\alpha)R$ and so by cancelling $[I : \mathcal{O}\alpha]_R = R$ in (16.4.4) we obtain (ii). To show (ii) $\Rightarrow$ (i), suppose that $\mathrm{Nm}_{B|F}(I) = [\mathcal{O} : I]_R$. Let $\alpha \in I$ be such that $\mathrm{Nm}_{B|F}(\alpha)$ has minimal valuation; then $\mathrm{Nm}_{B|F}(\alpha)$ generates $\mathrm{Nm}_{B|F}(I)$. By (16.4.4), cancelling on both sides $[I : \mathcal{O}\alpha]_R = R$, and since $\mathcal{O}\alpha \subseteq I$ we conclude $I = \mathcal{O}\alpha$. A similar argument holds on the right, proving (i) $\Leftrightarrow$ (iii). Finally, (iii) $\Leftrightarrow$ (iv) since $\mathrm{Nm}_{B|F}(\alpha) = \mathrm{nrd}(\alpha)^n$, and the same on the right. □

16.4.5. Recalling the proof of Proposition 16.4.3 and the definition of R-index, we always have the containment

$$\mathrm{Nm}_{B|F}(I) \supseteq [\mathcal{O}_{\mathsf{L}}(I) : I]_R$$

and the same on the right; by Proposition 16.4.3, equality is equivalent to I being locally principal.

To conclude this section, we suppose for its remainder that F is a local with valuation ring R or a global number field. Then the reduced norm is also related to the absolute norm, an absolute measure of size, as follows.

16.4.6. For a fractional ideal $\mathfrak{a}$ of R, we define the **absolute norm** $\mathsf{N}(\mathfrak{a})$ to be

$$\mathsf{N}(\mathfrak{a}) := [R : \mathfrak{a}]_{\mathbb{Z}} < \infty \tag{16.4.7}$$

the index taken as abelian groups, recalling Example 9.6.6. Then

$$\mathsf{N}(\mathfrak{a}) = |\mathrm{Nm}_{F/\mathbb{Q}}(\mathfrak{a})|$$

and if $\mathfrak{a} \subseteq R$ then

$$\mathsf{N}(\mathfrak{a}) = \#(R/\mathfrak{a}),$$

so this norm is also called the **counting norm**.

We extend this definition to elements $a \in F^\times$ by defining $\mathsf{N}(a) := \mathsf{N}(aR)$.

16.4.8. Similarly, if $I \subseteq B$ is a locally principal R-lattice, we define the **absolute norm** of I to be

$$\mathsf{N}(I) := [\mathcal{O}_{\mathsf{L}}(I) : I]_{\mathbb{Z}} = [\mathcal{O}_{\mathsf{R}}(I) : I]_{\mathbb{Z}}, \tag{16.4.9}$$

the latter equality by taking $R = \mathbb{Z}$ in Proposition 16.4.3. If I is integral then

$$\mathsf{N}(I) = \#(\mathcal{O}_{\mathsf{L}}(I)/I) = \#(\mathcal{O}_{\mathsf{R}}(I)/I).$$

The absolute norm of I is compatible with the absolute norm on R via

$$\mathsf{N}(I) = [\mathcal{O}_{\mathsf{R}}(I) : I]_{\mathbb{Z}} = [R : \mathrm{Nm}_{B|F}(I)]_{\mathbb{Z}} = \mathsf{N}(\mathrm{Nm}_{B|F}(I));$$

and if B is simple with $\dim_F B = n^2$ then

$$\mathsf{N}(I) = \mathsf{N}(\mathrm{Nm}_{B|F}(I)) = \mathsf{N}(\mathrm{nrd}(I))^n. \tag{16.4.10}$$

Remark 16.4.11. The absolute norm may also be defined for a global function field, but there is no canonical 'ring of integers' as above.

16.5 Invertible lattices

We are now in a position to investigate the class of invertible lattices. Let $I \subseteq B$ be an R-lattice.

Definition 16.5.1. I is **invertible** if there exists an R-lattice $I' \subseteq B$ that is a **(two-sided) inverse** to I, i.e.

$$II' = O_{\mathsf{L}}(I) = O_{\mathsf{R}}(I') \text{ and } I'I = O_{\mathsf{L}}(I') = O_{\mathsf{R}}(I). \tag{16.5.2}$$

In particular, both of the products in (16.5.2) are compatible.

16.5.3. If I, J are invertible lattices and I is compatible with J, then IJ is invertible (Exercise 16.10).

16.5.4. If I is a principal lattice, then I is invertible: if $I = O\alpha$ with $\alpha \in B^\times$ and $O = O_{\mathsf{L}}(I)$, then $I' = \alpha^{-1}O$ has

$$II' = (O\alpha)(\alpha^{-1}O) = O(\alpha\alpha^{-1})O = OO = O$$

so I' is a right inverse, and

$$I'I = (\alpha^{-1}O)(O\alpha) = \alpha^{-1}O\alpha = O_{\mathsf{R}}(I)$$

so I' is also a left inverse.

A candidate for the inverse presents itself quite naturally. If $II' = O_{\mathsf{L}}(I)$ and $I'I = O_{\mathsf{R}}(I)$, then $II'I = I$.

Definition 16.5.5. We define the **quasi-inverse** of I as

$$I^{-1} := \{\alpha \in B : I\alpha I \subseteq I\}. \tag{16.5.6}$$

Lemma 16.5.7. *The following statements hold.*

(a) *The quasi-inverse I^{-1} is an R-lattice and*

$$II^{-1}I \subseteq I.$$

(b) *If O is an R-order, then $O^{-1} = O$.*

Proof. Statement (a) follows as in the proof of Lemma 10.2.7, and the inclusion is by the definition of I^{-1}. For statement (b), if $\alpha \in O$, then $O\alpha O \subseteq O$ since O is an order; conversely, if $O\alpha O \subseteq O$, then taking $1 \in O$ on left and right we conclude $\alpha \in O$. ◻

We now consider the quasi-inverse as an inverse.

Proposition 16.5.8. *The following are equivalent:*

(i) I^{-1} *is a (two-sided) inverse for* I*;*
(ii) $I^{-1}I = \mathcal{O}_{\mathsf{R}}(I)$ *and* $II^{-1} = \mathcal{O}_{\mathsf{L}}(I)$*;*
(iii) I *is invertible;*
(iv) *There is a compatible product* $II^{-1}I = I$ *and both* $1 \in II^{-1}$ *and* $1 \in I^{-1}I$*.*

Proof. The implication (i) ⇒ (ii) is clear. For (ii) ⇒ (i), we need to check the compatibility of the product: but since $I^{-1}I = \mathcal{O}_{\mathsf{R}}(I)$ we have $\mathcal{O}_{\mathsf{L}}(I^{-1}) \subseteq \mathcal{O}_{\mathsf{R}}(I)$, and from the other direction we have the other containment, so these are equal.

The implication (i) ⇒ (iii) is clear. For (iii) ⇒ (i), suppose that I' is an inverse to I. Then $I = II'I$ so $I' \subseteq I^{-1}$ by definition. Therefore $I \subseteq II^{-1}I \subseteq I$ and equality holds throughout. Multiplying by I' on the left and right then gives

$$I^{-1} = (I'I)I^{-1}(II') = I'II' = I'.$$

Again the implication (i) ⇒ (iv) is immediate. To prove (iv) ⇒ (ii), we need to show that $II^{-1} = \mathcal{O}_{\mathsf{L}}(I)$ and $I^{-1}I = \mathcal{O}_{\mathsf{R}}(I)$; we show the former. By compatibility, $\mathcal{O}_{\mathsf{R}}(I^{-1}) = \mathcal{O}_{\mathsf{L}}(I) = \mathcal{O}$. If $II^{-1} = J$ then $J = II^{-1} = \mathcal{O}(II^{-1})\mathcal{O} = \mathcal{O}J\mathcal{O}$, so $J \subseteq \mathcal{O}$ is a two-sided ideal of $\mathcal{O}$ containing 1 hence $J = \mathcal{O}$. □

Invertibility is a local property, as one might expect.

Lemma 16.5.9. *I is invertible if and only $I_{(\mathfrak{p})}$ is invertible for all primes $\mathfrak{p}$.*

Proof. We employ Proposition 16.5.8(iv): We have $II^{-1}I = I$ if and only if

$$(II^{-1}I)_{(\mathfrak{p})} = I_{(\mathfrak{p})}(I^{-1})_{(\mathfrak{p})}I_{(\mathfrak{p})} = I_{(\mathfrak{p})}$$

for all primes $\mathfrak{p}$ and e.g. $1 \in II^{-1}$ if and only if $1 \in I_{(\mathfrak{p})}I_{(\mathfrak{p})}^{-1}$. □

Corollary 16.5.10. *If I is locally principal, then I is invertible.*

Proof. Combine 16.5.4 with Lemma 16.5.9. □

A compatible product with an invertible lattice respects taking left (and right) orders, as follows.

Lemma 16.5.11. *If I is compatible with J and J is invertible, then $\mathcal{O}_{\mathsf{L}}(IJ) = \mathcal{O}_{\mathsf{L}}(I)$.*

Proof. We always have $\mathcal{O}_{\mathsf{L}}(I) \subseteq \mathcal{O}_{\mathsf{L}}(IJ)$ (even without J invertible). To show the other containment, suppose that $\alpha \in \mathcal{O}_{\mathsf{L}}(IJ)$, so that $\alpha IJ \subseteq IJ$. Multiplying by J^{-1}, we conclude $\alpha I \subseteq I$ and $\alpha \in \mathcal{O}_{\mathsf{L}}(I)$. □

Finally, not every lattice is invertible, and it is helpful to have counterexamples at hand (see also Exercise 16.12).

Example 16.5.12. Let $p \in \mathbb{Z}$ be prime. Let $B := \left(\dfrac{p,p}{\mathbb{Q}}\right)$ and

$$\begin{aligned} O &:= \mathbb{Z} \oplus p\mathbb{Z}i \oplus p\mathbb{Z}j \oplus \mathbb{Z}ij \\ I &:= p^2\mathbb{Z} \oplus \mathbb{Z}i \oplus \mathbb{Z}j \oplus \mathbb{Z}ij. \end{aligned}$$

Then $O \subset B$ is an order and $O_{\mathsf{L}}(I) = O_{\mathsf{R}}(I) = O$. We compute that

$$I^{-1} = p\mathbb{Z} \oplus \mathbb{Z}i \oplus \mathbb{Z}j \oplus \mathbb{Z}ij \tag{16.5.13}$$

and

$$O_{\mathsf{L}}(I^{-1}) = O_{\mathsf{R}}(I^{-1}) = \mathbb{Z} + \mathbb{Z}i + \mathbb{Z}j + \frac{1}{p}\mathbb{Z}ij = \mathbb{Z} + \frac{1}{p}O; \tag{16.5.14}$$

so in the product

$$II^{-1} = I^{-1}I = p\mathbb{Z} \oplus p\mathbb{Z}i \oplus p\mathbb{Z}j \oplus \mathbb{Z}ij \subsetneq O \tag{16.5.15}$$

we see I is not invertible and the product is not compatible.

Seen a different way, we have $\overline{I} = I$ and in the compatible product

$$I^2 = I\overline{I} = \overline{I}I = p\mathbb{Z} \oplus p\mathbb{Z}i \oplus p\mathbb{Z}j \oplus \mathbb{Z}ij \tag{16.5.16}$$

we have $i, j \in O_{\mathsf{L}}(I^2) = O_{\mathsf{R}}(I^2)$ but $i, j \notin O$; therefore, I is not invertible by Lemma 16.5.11. Indeed,

$$O_{\mathsf{L}}(I^2) = O_{\mathsf{R}}(I^2) = \frac{1}{p}I^2 = \mathbb{Z} + \frac{1}{p}O.$$

Finally, it will convenient to consider invertibility in the context of ideals, labelling left and right orders as follows.

Definition 16.5.17. Let $O, O' \subseteq B$ be R-orders and let I be a fractional O, O'-ideal. We say I is **invertible** if I is invertible as a lattice and I is sated (i.e., $O = O_{\mathsf{L}}(I)$ and $O' = O_{\mathsf{R}}(I)$).

16.5.18. The condition that I is *sated* in Definition 16.5.17 is important: we must be careful to work over left and right orders and not some smaller order. Indeed, if I is invertible as an R-lattice then it is invertible as a fractional $O_{\mathsf{L}}(I), O_{\mathsf{R}}(I)$-ideal, but *not* for any strictly smaller orders. If I' is an R-lattice and $II' = O$ for some $O \subseteq O_{\mathsf{L}}(I)$, then multiplying on both sides on the left by $O_{\mathsf{L}}(I)$ gives

$$O = II' = O_{\mathsf{L}}(I)II' = O_{\mathsf{L}}(I)O = O_{\mathsf{L}}(I)$$

and the same on the right. In other words, if we are going to call out an invertible fractional ideal by labelling actions on left and right, then we require these labels to be the actual orders that make the inverse work.

Remark 16.5.19. Equation 16.1.1 suggested the ‘real issue’ with noninvertible modules for *quadratic* orders: as an abelian group,

$$\mathfrak{f} = f\mathbb{Z} + f\sqrt{d}\mathbb{Z} = f \cdot S(d),$$

so $\mathfrak{f}$ is principal and hence certainly invertible as an ideal of $S(d)$—but not as an ideal of the smaller order $S(d)$. More generally, if $\mathfrak{a} \subset K = \mathbb{Q}(\sqrt{d})$ is a lattice in K (free $\mathbb{Z}$-module of rank 2), we define its **multiplicator ring** as

$$S(\mathfrak{a}) := \{x \in K : x\mathfrak{a} \subseteq \mathfrak{a}\};$$

the ring $S(\mathfrak{a})$ is an order of K and so is also called the **order** of $\mathfrak{a}$. In the example above, $S(\mathfrak{f}) = S(f(\mathbb{Z}+\sqrt{d_K}\mathbb{Z})) = S(d_K) \supsetneq S(d)$. It turns out that every lattice in K is invertible as an ideal of its multiplicator ring [Cox89, Proposition 7.4], and this statement plays an important role in the theory of complex multiplication. (Sometimes, an ideal $\mathfrak{a} \subseteq S$ is called **proper** or **regular** if $S = S(\mathfrak{a})$; both terms are overloaded in mathematics, so we will mostly resist this notion.)

Unfortunately, unlike the quadratic case, not every lattice $I \subset B$ is projective as a left module over its left order (or the same on the right): this is necessary, but not sufficient. In Chapter 24, we classify orders O with the property that every lattice I having $O_{\mathsf{L}}(I) = O$ is projective as an O-module: they are the *Gorenstein* orders.

Remark 16.5.20. Invertible lattices give rise to a *Morita equivalence* between their corresponding left and right orders: see Remark 7.2.20.

16.6 Invertibility with a standard involution

In section 16.6, we follow Kaplansky [Kap69], considering invertibility in the presence of a standard involution. The main result of this chapter is as follows.

Main Theorem 16.6.1. *Let R be a Dedekind domain with field of fractions F, and let B be a finite-dimensional F-algebra with a standard involution. Then an R-lattice I is invertible if and only if I is locally principal.*

Remark 16.6.2. We can relax the hypothesis that R is a Dedekind domain and instead work with a *Prüfer domain*, a generalization of Dedekind domains to the non-noetherian context.

We have already seen (Corollary 16.5.10) that the implication ($\Rightarrow$) in Main Theorem 16.6.1 holds without the hypothesis of a standard involution; the reverse implication is the topic of this section. This implication is not in general true if this hypothesis is removed (but is true again when B is commutative); see Exercise 16.18(a).

Remark 16.6.3. The provenance of the hypothesis that R is a Dedekind domain is the following: if $\mathfrak{a} \subset R$ is not invertible as an R-module, and $O \subset B$ is an R-order, then $\mathfrak{a}O$ is not invertible as an R-lattice. To make the simplest kind of arguments here, we would like for all (nonzero) ideals $\mathfrak{a} \subseteq R$ to be invertible, and this is equivalent to the requirement that R is a Dedekind domain (see section 9.2).

Throughout this section, let R be a Dedekind domain with field of fractions F, let B be a finite-dimensional F-algebra, and let $I \subset B$ be an R-lattice. The following concept will be useful in this section.

Definition 16.6.4. We say I is a **semi-order** if $1 \in I$ and $\operatorname{nrd}(I) \subseteq R$.

(For a semi-order I, we necessarily have $\operatorname{nrd}(I) = R$ since $1 \in I$.)

Lemma 16.6.5. *An R-lattice I is a semi-order if and only if $1 \in I$ and every $\alpha \in I$ is integral over R.*

Proof. We have that $\alpha \in I$ is integral over R if and only if $\operatorname{trd}(\alpha) \in R$ and $\operatorname{nrd}(\alpha) \in R$ (by Corollary 10.3.6, since R is integrally closed) if and only if $\operatorname{nrd}(\alpha) \in R$ and $\operatorname{nrd}(\alpha+1) = \operatorname{nrd}(\alpha) + \operatorname{trd}(\alpha) + 1 \in R$. □

In particular, Lemma 16.6.5 implies that an order is a semi-order (by Corollary 10.3.3); we will see that semi-orders behave enough like orders that we can deduce local principality from their structure.

16.6.6. Let $\overline{I} := \{\overline{\alpha} : \alpha \in I\}$. Then $\overline{I}$ is an R-lattice in B. If I, J are R-lattices then $\overline{IJ} = \overline{J}\,\overline{I}$ (even if this product is not compatible).

If I is a semi-order, then $\overline{I} = I$ (Exercise 16.15). In particular, if O is an R-order then $\overline{O} = O$.

Lemma 16.6.7. *We have $O_{\mathsf{L}}(I) = O_{\mathsf{R}}(\overline{I})$ and $O_{\mathsf{R}}(I) = O_{\mathsf{L}}(\overline{I})$.*

Proof. We have $\alpha \in O_{\mathsf{L}}(I)$ if and only if $\alpha I \subseteq I$ if and only if $\overline{\alpha I} = \overline{I}\,\overline{\alpha} \subseteq \overline{I}$ if and only if $\overline{\alpha} \in O_{\mathsf{R}}(\overline{I})$ if and only if $\alpha \in \overline{O_{\mathsf{R}}(\overline{I})} = O_{\mathsf{R}}(\overline{I})$. □

Corollary 16.6.8. *If I is a semi-order, then $O_{\mathsf{L}}(I) = O_{\mathsf{R}}(I)$.*

Proof. Apply Lemma 16.6.7 with $\overline{I} = I$. □

By Lemma 16.6.7, the standard involution gives a bijection between the set of lattices I with $O_{\mathsf{L}}(I) = O$ and the set of lattices with $O_{\mathsf{R}}(I) = O$.

16.6.9. Suppose that R is a DVR (e.g., a localization of R at a prime ideal $\mathfrak{p}$). We will show how to reduce the proof of Main Theorem 16.6.1 to that of a semi-order.

Since R is a DVR, the fractional R-ideal $\operatorname{nrd}(I) \subseteq R$ is principal, generated by an element with minimal valuation: let $\alpha \in I$ achieve this minimum reduced norm. Then the R-lattice $J = \alpha^{-1}I$ now satisfies $1 \in J$ and $\operatorname{nrd}(J) = R$. Thus J is a semi-order, and J is (locally) principal if and only if I is (locally) principal.

16.6.1

Proof of Main Theorem. The proof is due to Kaplansky [Kap69, Theorem 2]. The statement is local; localizing, we may suppose R is a DVR. By 16.6.9, we reduce to the case where I is a semi-order. In particular, we have $1 \in I$. Let $\alpha_1, \dots, \alpha_n$ be an R-basis for I.

We claim that

$$I^{n+1} = I^n \tag{16.6.10}$$

Since $1 \in I$, we have $I^n \subseteq I^{n+1}$. It suffices then to prove that a product of $n+1$ basis elements of I lies in I^n. By the pigeonhole principle, there must be a repeated term α_i among them. We recall the formula (4.2.16)

$$\alpha\beta + \beta\alpha = \operatorname{trd}(\beta)\alpha + \operatorname{trd}(\alpha)\beta - \operatorname{trd}(\alpha\overline{\beta}) \tag{16.6.11}$$

for all $\alpha, \beta \in B$. We can use this relation to "push" the second instance of the repeated element until it meets with its mate, at the expense of terms lying in I^n. More precisely, in the R-module I^2/I, by (16.6.11),

$$\alpha_i\alpha_j \equiv -\alpha_j\alpha_i \pmod{I}$$

for all i, j; it follows that in I^{n+1}/I^n,

$$\mu(\alpha_i\alpha_j)\nu \equiv -\mu(\alpha_j\alpha_i)\nu \pmod{I^n}$$

for all μ, ν appropriate products of basis elements. Therefore we may suppose that the repetition α_i^2 is adjacent; but then α_i satisfies a quadratic equation and $\alpha_i^2 = \operatorname{trd}(\alpha_i)\alpha_i - \operatorname{nrd}(\alpha_i) \in I$, so in fact the product belongs to I^n, and the claim follows.

Now suppose I is invertible; we wish to show that I is principal. From the equality $I^{n+1} = I^n$, we multiply both sides of this equation by $(I^{-1})^n$ and obtain $I = \mathcal{O} = \mathcal{O}_{\mathsf{L}}(I) = \mathcal{O}_{\mathsf{R}}(I)$. In particular, I is principal, generated by 1. □

The above proof has the following immediate corollary.

Corollary 16.6.12. *An R-lattice I is an R-order if and only if $1 \in I$, every element of I is integral, and I is invertible. In particular, an invertible semi-order is an order.*

We conclude with two consequences.

16.6.13. Let I, J be invertible R-lattices such that I is compatible with J. Then $\operatorname{nrd}(IJ) = \operatorname{nrd}(I)\operatorname{nrd}(J)$, since it is enough to check this locally, and locally both I and J are principal and we have proved the statement in this case (Lemma 16.3.7).

16.6.14. In the presence of a standard involution, we can write the inverse in another way: if I is invertible, then

$$\overline{I}I = \operatorname{nrd}(I)\mathcal{O}_{\mathsf{R}}(I) \quad \text{and} \quad I\overline{I} = \operatorname{nrd}(I)\mathcal{O}_{\mathsf{L}}(I)$$

by checking these statements locally (where they follow immediately by computing the norm on a local generator). Since $\operatorname{nrd}(I)$ is a fractional R-ideal and thus invertible (R is a Dedekind domain), it follows that if I is invertible, then

$$I^{-1} = \overline{I}\operatorname{nrd}(I)^{-1}.$$

In view of 16.6.14, the following important proposition is natural.

Proposition 16.6.15. *Let B be a quaternion algebra over F and let $I \subset B$ be an R-lattice. Then the following statements hold.*

(a) *We have $I\overline{I} = \operatorname{nrd}(I)\mathcal{O}$, where $\mathcal{O} \subseteq B$ an R-order satisfying $\mathcal{O}_{\mathsf{L}}(I) \subseteq \mathcal{O}$, and similarly $\overline{I}I = \operatorname{nrd}(I)\mathcal{O}'$ with $\mathcal{O}_{\mathsf{R}}(I) \subseteq \mathcal{O}'$.*

(b) *If either $\mathcal{O}_{\mathsf{L}}(I)$ or $\mathcal{O}_{\mathsf{R}}(I)$ is maximal, then I is invertible, and both $\mathcal{O}_{\mathsf{L}}(I)$ and $\mathcal{O}_{\mathsf{R}}(I)$ are maximal.*

Proof. We follow Kaplansky [Kap69, Theorems 6–7]. We again may suppose R is a DVR and I is a semi-order, so $1 \in I$ and $\operatorname{nrd}(I) = R$; and $\overline{I} = I$.

First we prove (a). We need to show that I^2 is an order. We showed in (16.6.10) (without extra hypothesis) that $I^3 = I^4$; with B a quaternion algebra, we will improve this to $I^2 = I^3$, whence $(I^2)^2 = I^4 = I^2$ and consequently I^2 is closed under multiplication and hence an R-order.

Let $J = I^3$; then $J^2 = (I^3)^2 = I^6 = I^3 = J$, so J is an R-order.

Let $\mathfrak{p}$ be the maximal ideal of R and consider the 4-dimensional algebra $J/\mathfrak{p}J$ over $k = R/\mathfrak{p}$. Then $I/\mathfrak{p}I \subseteq J/\mathfrak{p}J$ is a k-subspace containing 1. If $(I/\mathfrak{p}I)^2 = I/\mathfrak{p}I$, then by dimensions we contradict $(I/\mathfrak{p}I)^3 = J/\mathfrak{p}J$; therefore $(I/\mathfrak{p}I)^2 \supsetneq I/\mathfrak{p}I$. If $\dim_k(I/\mathfrak{p}I) \leq 2$, then $I/\mathfrak{p}I$ is a proper k-subalgebra, impossible. Thus $\dim_k(I/\mathfrak{p}I) \geq 3$ and $\dim_k(I/\mathfrak{p}I)^2 \geq 4$, and so $(I/\mathfrak{p}I)^2 = J/\mathfrak{p}J$. By Nakayama's lemma, it follows that $I^2 = J = I^3$. The containments follow directly, e.g. $O_{\mathsf{L}}(I) \subseteq O_{\mathsf{L}}(I\overline{I}) = O_{\mathsf{L}}(O) = O$.

For part (b), applying part (a) we have $O = O_{\mathsf{L}}(I)$ by maximality and the same on the right. But since I is a semi-order, from Corollary 16.6.8 we have $I\overline{I} = O_{\mathsf{L}}(I) = O_{\mathsf{R}}(I) = \overline{I}I$ so by definition, I has inverse $\overline{I}$. □

16.7 One-sided invertibility

In this section, we pause to consider one-sided notions of invertibility. We refresh our notation, recalling that R is a Dedekind domain with $F = \operatorname{Frac} R$ and B is a finite-dimensional algebra over F with $I \subseteq B$ an R-lattice.

Definition 16.7.1. I is **right invertible** if there exists an R-lattice $I' \subseteq B$, a **right inverse**, such that the product II' is compatible and $II' = O_{\mathsf{L}}(I)$.

A right fractional O-ideal I is **right invertible** if I is right invertible and sated (viz. 16.5.18).

We similarly define **left invertible** and **left inverse**. Applying the same reasoning as in Lemma 16.5.9, we see that one-sided invertibility is a local property.

Remark 16.7.2. For rings, the (left or) right inverse of an *element* need not be unique even though a two-sided inverse is necessarily unique. Similarly, left invertibility does not imply right invertibility for lattices in general, and so the one-sided notions can be a bit slippery: see Exercise 16.18(b).

Remark 16.7.3. The compatibility condition in invertibility is important to avoid trivialities. Consider Example 16.3.6: we have $IJ = \mathrm{M}_2(R) = O_{\mathsf{L}}(I)$, and if we let $J = \begin{pmatrix} bR & bR \\ R & R \end{pmatrix}$ for *any* nonzero $b \in R$, the equality $IJ = \mathrm{M}_2(R)$ remains true. Not every author requires compatibility in the definition of (sided) invertibility.

A natural candidate for the right inverse presents itself: if $II' = O_{\mathsf{L}}(I)$, then I' maps I into $O_{\mathsf{L}}(I)$ on the right. We recall the definition of the colon lattices (Definition 15.6.11). Let $I' := (O_{\mathsf{L}}(I) : I)_{\mathsf{R}}$. Then $II' \subseteq O_{\mathsf{L}}(I)$ by definition; however, in general equality need not hold and the product need not be compatible. Similarly, since $II^{-1}I \subseteq I$ we have $II^{-1} \subseteq O_{\mathsf{L}}(I)$, but again equality need not hold.

The sided version of Proposition 16.5.8 also holds.

Proposition 16.7.4. *The following are equivalent:*

(i) I^{-1} *is a right inverse for* I*;*
(ii) I *is right invertible;*
(iii) *There is a compatible product* $II^{-1}I = I$ *and* $1 \in II^{-1}$.

Similar equivalences hold on the left.

Proof. This is just a sided restriction of the proof of Proposition 16.5.8. For example, to show (ii) ⇒ (i), we always have $II^{-1}I \subseteq I$ so $II^{-1} \subseteq \mathcal{O}_{\mathsf{L}}(I)$; if I' is a right inverse to I, then $II'I = \mathcal{O}_{\mathsf{L}}(I)I = I$ and $I' \subseteq I^{-1}$, and therefore $II^{-1} \supseteq II' = \mathcal{O}_{\mathsf{L}}(I)$. Therefore $II^{-1} = \mathcal{O}_{\mathsf{L}}(I)$ and I^{-1} is a right inverse for I. □

Returning to the setting of the previous section, however, we can show that the one-sided notions of invertibility are equivalent to the two-sided notion.

Lemma 16.7.5. *Suppose* B *has a standard involution. Then an* R*-lattice* I *is left invertible if and only if* I *is right invertible if and only if* I *is invertible.*

Proof. We will show that if I is right invertible then I is left invertible; the other implications follow similarly. By localizing, we reduce to the case where R is a DVR. By the results of 16.6.9, we may suppose that I is a semi-order, so that $\mathcal{O}_{\mathsf{L}}(I) = \mathcal{O}_{\mathsf{R}}(I) = \mathcal{O}$ and $I = \overline{I}$. Suppose $II' = \mathcal{O}$. Then $\overline{I'}I = \overline{\mathcal{O}} = \mathcal{O}$, and $\overline{I'}$ is compatible with I since

$$\mathcal{O} = \mathcal{O}_{\mathsf{R}}(I) = \mathcal{O}_{\mathsf{L}}(I') = \mathcal{O}_{\mathsf{R}}(\overline{I'})$$

as desired. □

Corollary 16.7.6. *Suppose* R *is a Dedekind domain and that* B *has a standard involution. Then an* R*-lattice* I *is right invertible with* $II' = \mathcal{O}_{\mathsf{L}}(I)$ *if and only if* $I' = (\mathcal{O}_{\mathsf{L}}(I) : I)_{\mathsf{R}} = I^{-1}$.

A similar statement holds for the left inverse; in particular, this shows that a right inverse is necessarily unique.

Proof. The implication (⇒) is immediate, so we prove (⇐). Let $\mathcal{O} = \mathcal{O}_{\mathsf{L}}(I)$. Then

$$\mathcal{O} = II' \subseteq I(\mathcal{O} : I)_{\mathsf{R}} \subseteq \mathcal{O}$$

so equality must hold, and $II' = I(\mathcal{O} : I)_{\mathsf{R}}$. By 16.7.5, I is invertible, and multiplying both sides by I^{-1} gives $I' = (\mathcal{O} : I)_{\mathsf{R}}$. □

We collect the results of this section in the following theorem.

Main Theorem 16.7.7. *Let* R *be a Dedekind domain with* $F = \operatorname{Frac} R$*, let* B *be a quaternion algebra over* F*, and let* $I \subseteq B$ *be an* R*-lattice. Then the following are equivalent:*

(i) I *is locally principal;*
(ii) I *is invertible;*
(iii) I *is left invertible;*
(iii′) I *is right invertible;*

(iv) $\operatorname{nrd}(I)^2 = [O_{\mathsf{L}}(I) : I]_R$; *and*
(iv′) $\operatorname{nrd}(I)^2 = [O_{\mathsf{R}}(I) : I]_R$.

Proof. Main Theorem 16.6.1 proves (i) ⇔ (ii). For the equivalence (ii) ⇔ (iii) ⇔ (iii′), apply Lemma 16.7.5. Finally, the equivalence (i) ⇔ (iv) ⇔ (iv′) is supplied by Proposition 16.4.3. □

16.8 Invertibility and the codifferent

To conclude this chapter, we pick up a remaining thread concerning the (co)different.

Definition 16.8.1. We define the **different** of O to be the quasi-inverse of the codifferent:

$$\operatorname{diff}(O) := \operatorname{codiff}(O)^{-1} = \{\alpha \in B : O^\sharp \alpha O^\sharp \subseteq O^\sharp\}.$$

Lemma 16.8.2. *The different* $\operatorname{diff}(O)$ *is an integral two-sided* O*-ideal.*

Proof. By Lemma 15.6.16, we have $OO^\sharp O = O^\sharp$ and so if $\alpha \in \operatorname{diff}(O)$ then $O^\sharp(O\alpha O)O^\sharp = O^\sharp\alpha O^\sharp = O^\sharp$ and $\operatorname{diff}(O)$ is a two-sided O-ideal. To prove that $\operatorname{diff}(O) \subseteq O$, referring to Lemma 15.6.2, starting with $O^\sharp\alpha O^\sharp \subseteq O^\sharp$ taking $1 \in O^\sharp$ we have $\alpha O^\sharp \subseteq O^\sharp$ so $(\alpha O^\sharp)^\sharp = (O^\sharp)^\sharp\alpha^{-1} \supseteq (O^\sharp)^\sharp$. By Lemma 15.6.5, we have $(O^\sharp)^\sharp = O$, so $O\alpha^{-1} \supseteq O$, so $O\alpha \subseteq O$ and again taking 1 we get $\alpha \in O$. □

16.8.3. If $\operatorname{codiff}(O)$ is locally principal (section 16.2), then so is $\operatorname{diff}(O)$, and by Proposition 16.4.3 we have

$$\operatorname{Nm}_{B|F}(\operatorname{diff}(O)) = [O : \operatorname{diff}(O)]_R = [\operatorname{codiff}(O) : O]_R = \operatorname{disc}(O);$$

so when further B is a quaternion algebra, we have

$$\operatorname{nrd}(\operatorname{diff}(O)) = \operatorname{discrd}(O). \tag{16.8.4}$$

Invertibility of ideals is detected by the (co)different [Fad65, Proposition 24.1].

Proposition 16.8.5. *If* $\operatorname{codiff}(O)$ *is right invertible, then all sated left fractional* O*-ideals are right invertible. Similarly, if* $\operatorname{codiff}(O)$ *is left invertible, then all sated right fractional* O*-ideals are left invertible.*

Proof. To get started, we refresh a few things: by Corollary 15.6.13, we have $(II^\sharp)^\sharp = O_{\mathsf{L}}(I) = O$. The product $II^\sharp$ is compatible by Proposition 15.6.6. By Lemma 15.6.5 we have $II^\sharp = O^\sharp = \operatorname{codiff}(O)$.

Now by hypothesis of invertibility, $O^\sharp(O^\sharp)^{-1} = O_{\mathsf{L}}(O^\sharp) = O$ is a compatible product. Therefore the product $I^\sharp(O^\sharp)^{-1}$ is compatible, and

$$I(I^\sharp(O^\sharp)^{-1}) = (II^\sharp)(O^\sharp)^{-1} = O^\sharp(O^\sharp)^{-1} = O. \tag{16.8.6}$$

A similar argument holds on the right. □

We have the following corollary of Proposition 16.8.5, phrased in terms of the different.

Corollary 16.8.7. *Suppose that B has a standard involution. Then the following are equivalent:*

(i) $\operatorname{codiff}(O)$ *is invertible;*
(ii) $\operatorname{diff}(O)$ *is invertible;*
(iii) *All sated left fractional O-ideals I are invertible, with inverse $I^{-1} = I^\sharp \operatorname{diff}(O)$; and*
(iii′) *All sated right fractional O-ideals I are invertible, with inverse $I^{-1} = \operatorname{diff}(O) I^\sharp$.*

Proof. Combine Proposition 16.8.5 and (16.8.6) with Lemma 16.7.5 and Corollary 16.7.6. □

We conclude with a criterion to determine invertibility; it is not used in the sequel.

Proposition 16.8.8. (Brandt's invertibility criterion). *Let $I \subseteq B$ be an R-lattice. Then I is invertible if and only if*

$$\operatorname{nrd}(I^\sharp)\operatorname{discrd}(I) \subseteq \operatorname{nrd}(I).$$

Proof. See Kaplansky [Kap69, Theorem 10] or Brzezinski [Brz82, Theorem 3.4]. □

Exercises

Unless otherwise specified, throughout these exercises let R be a Dedekind domain with field of fractions F, let B be a finite-dimensional F-algebra, and let $I \subseteq B$ be an R-lattice.

1. Let $d \in \mathbb{Z}$ be a nonsquare discriminant, and let $S(d) = \mathbb{Z}[(d + \sqrt{d})/2]$ be the quadratic ring of discriminant d.
   (a) Suppose that $d = d_K f^2$ with $f > 1$. Show that the ideal $(f, \sqrt{d})$ of $S(d)$ is not invertible.
   (b) Consider $d = -12$, and $S = S(-12) = \mathbb{Z}[\sqrt{-3}]$. Show that every invertible ideal of S is principal (so S has class number 1), but that S is not a PID.

▹2. Show that if $I = O_{\mathsf{L}}(I)\alpha$ with $\alpha \in B^\times$, then $O_{\mathsf{R}}(I) = \alpha^{-1} O_{\mathsf{L}}(I)\alpha$.

▹3. Show that if J is an R-lattice in B and $\mu \in B^\times$, then $\mu J = J$ if and only if $\mu \in O_{\mathsf{L}}(J)^\times$.

▹4. Show that if $\alpha \in B$ then $\operatorname{nrd}(\alpha I) = \operatorname{nrd}(\alpha)\operatorname{nrd}(I)$. Conclude that if I is a principal R-lattice, generated by $\alpha \in I$, then $\operatorname{nrd}(I) = \operatorname{nrd}(\alpha)R$.

5. Let $\alpha_1, \dots, \alpha_n$ generate I as an R-module. Give an explicit example where $\operatorname{nrd}(I)$ is *not* generated by $\operatorname{nrd}(\alpha_i)$ (cf. Lemma 16.3.2). Moreover, show that for an R-lattice I, there exists a set of R-module generators α_i such that $\operatorname{nrd}(I)$ is in fact generated by $\operatorname{nrd}(\alpha_i)$.

▹6. Suppose that R is a Dedekind domain, and let $O \subseteq B$ be an R-order. Let I be a locally principal right fractional O-ideal. Show that I can be generated as a right O-ideal by two elements, and in fact for $a \in \operatorname{nrd}(I)$ nonzero we can write $I = aO + \beta O$ with $\beta \in B^\times$.

7. Let F be a number field, let $R \subseteq F$ be a ($\mathbb{Z}$-)order, and let $\mathfrak{a} \subseteq R$ be a nonzero ideal. Show that $\mathfrak{a}$ is projective as an R-module if and only if $\mathfrak{a}$ is invertible if and only if $\mathfrak{a}$ is locally principal. [These are all automatic when R is a Dedekind domain 9.4.5.]
8. If $I, J \subseteq B$ are R-lattices with $I \subseteq J$, is it true that $I^{-1} \supseteq J^{-1}$?
9. Let $I, J, K \subseteq B$ be R-lattices. Show that

$$((I : J)_{\mathsf{L}} : K)_{\mathsf{R}} = ((I : K)_{\mathsf{R}} : J)_{\mathsf{L}}.$$

▹10. Let $I, J \subseteq B$ be R-lattices and suppose that I is compatible with J. Show that IJ is invertible (with $(IJ)^{-1} = J^{-1}I^{-1}$) if both I, J are invertible, but the converse need not hold.

11. Let $I, J \subseteq B$ be R-lattices, and suppose that J is invertible. Show that $(I : J)_{\mathsf{L}} = IJ^{-1}$ and $(I : J)_{\mathsf{R}} = J^{-1}I$.
12. Let p be prime, let $B = (p, p \mid \mathbb{Q})$, and let $O := \mathbb{Z}\langle i, j\rangle = \mathbb{Z} \oplus \mathbb{Z}i \oplus \mathbb{Z}j \oplus \mathbb{Z}ij$.

    (a) Let $I = \{\alpha \in O : p \mid \operatorname{nrd}(\alpha)\}$. Show that $I = p\mathbb{Z} \oplus \mathbb{Z}i \oplus \mathbb{Z}j \oplus \mathbb{Z}ij$.
    (b) Show $I = Oi + Oj$, that O is a two-sided O-ideal, and that $[O : I] = p$.
    (c) Show that $I_{(p)} \neq O_{(p)}\alpha$ for all $\alpha \in I_{(p)}$. *[Hint: show that if $\alpha \in I$, then $p^2 \mid [O : O\alpha]$.]*
    (d) Compute that $O_{\mathsf{L}}(I) = \mathbb{Z} + \mathbb{Z}i + \mathbb{Z}j + \mathbb{Z}(ij/p) \supsetneq O$, and that $I = O_{\mathsf{L}}(I)i = O_{\mathsf{L}}(I)j$.
    (e) Compute $\operatorname{codiff}(O)$ and $\operatorname{diff}(O)$ and show they are invertible.

    [Compare Lemurell [Lem2011, Remark 6.4].]

▹13. Let K be a separable quadratic field extension of F and let $I \subseteq K$ be an R-lattice. Let $O = O_{\mathsf{L}}(I) = O_{\mathsf{R}}(I)$.

    (a) Show that $I\overline{I} = \overline{I}I = \operatorname{nrd}(I)O$. *[Hint: argue as in Proposition 16.6.15.]*
    (b) Conclude that I is invertible as a O-module.

14. Show that if $I, J \subseteq B$ are locally principal (hence invertible) R-lattices, then

$$[I : J]_R = [J^{-1} : I^{-1}]_R.$$

▹15. Let B be an F-algebra with a standard involution $\overline{}$. Show that if I is a semi-order then $\overline{I} = I$.

16. Let R be a Dedekind domain with field of fractions F, let $K \supset F$ be a separable quadratic field extension and let S be an R-order in K. Let S_K be the integral closure of R in K.

    (a) Show that there exists a (unique) ideal $\mathfrak{f} = \mathfrak{f}(S) \subset S_K$ (called the **conductor**) such that $S = R + \mathfrak{f}S_K$.
    (b) Now let $\mathfrak{b} \subset K$ be a fractional S-ideal. Show that the following are equivalent:

        (i) $\mathfrak{b}$ is a locally principal S-ideal;

(ii) $\mathfrak{b}$ is invertible as a fractional S-ideal, i.e., there exists a fractional ideal $\mathfrak{b}^{-1}$ such that $\mathfrak{b}\mathfrak{b}^{-1} = S$ (necessarily $\mathfrak{b}^{-1} = (S : \mathfrak{b})$);
(iii) There exists $d \in K^\times$ such that $d\mathfrak{b} + \mathfrak{f} \cap S = S$; and
(iv) $\mathfrak{b}$ is proper, i.e., $S = O(\mathfrak{b}) = \{x \in K : x\mathfrak{b} \subseteq \mathfrak{b}\}$.

▹17. Let $O \subseteq B$ be an R-order.

(a) Let $\alpha \in B^\times$. Show that $I = O\alpha$ is a lattice with $O_\mathsf{L}(I) = O_\mathsf{R}(I) = O$ if and only if $\alpha \in B^\times$ and $O\alpha = \alpha O$. Conclude that the set of invertible two-sided principal lattices I with $O_\mathsf{L}(I) = O_\mathsf{R}(I) = O$ forms a group.
(b) Show that the normalizer of O,

$$N_{B^\times}(O) = \{\alpha \in B^\times : \alpha O \alpha^{-1} = O\}$$

is the group generated by $\alpha \in B^\times$ such that $O\alpha$ is a two-sided O-ideal.

18. The following example is due to Kaplansky [Kap69, pp. 220, 221]. Let R be a DVR with field of fractions F and maximal ideal $\mathfrak{p} = \pi R$.

(a) Consider the R-lattice

$$I = \begin{pmatrix} \pi R & \pi R & R \\ \pi R & \pi R & R \\ R & R & R \end{pmatrix} \subset B = \mathrm{M}_3(F)$$

Show that I is invertible but is not principal.
(b) Consider the R-lattice

$$I = \begin{pmatrix} \pi R & \pi R & R \\ \pi^2 R & \pi^2 R & R \\ R & R & R \end{pmatrix} \subset B = \mathrm{M}_3(F)$$

Show that I is left invertible but is not right invertible.

This chapter is licensed under the terms of the Creative Commons Attribution-NonCommercial 4.0 International License (http://creativecommons.org/licenses/by-nc/4.0/), which permits any noncommercial use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license and indicate if changes were made.

The images or other third party material in this chapter are included in the chapter's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the chapter's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.

Chapter 17

Classes of quaternion ideals

Having investigated the structure of lattices and ideals in Chapter 16, we now turn to the study of their isomorphism classes.

17.1 ▹ Ideal classes

For motivation, let K be a quadratic number field and $S \subseteq K$ an order. We say that two invertible fractional ideals $\mathfrak{a}, \mathfrak{b} \subset K$ of S are **in the same class**, and write $\mathfrak{a} \sim \mathfrak{b}$, if there exists $c \in K^\times$ such that $c\mathfrak{a} = \mathfrak{b}$; we denote the class of a fractional ideal $\mathfrak{a}$ as $[\mathfrak{a}]$. We have $\mathfrak{a} \sim \mathfrak{b}$ if and only if $\mathfrak{a}$ and $\mathfrak{b}$ are isomorphic as S-modules. The set $\operatorname{Cl} S$ of invertible fractional ideals is a group under multiplication, measuring the failure of S to be a PID. The class group $\operatorname{Cl} S$ is a finite abelian group, by Minkowski's *geometry of numbers*: every class in $\operatorname{Cl} S$ is represented by an integral ideal $\mathfrak{a} \subseteq S$ whose absolute norm is bounded (depending on S, but independent of the class), and there are only finitely many such ideals. For an introduction to orders in quadratic fields and their class numbers, with further connections to quadratic forms, see Cox [Cox89, §7].

The first treatment of isomorphism classes of quaternion ideals was given by Brandt [Bra28]. Let B be a quaternion algebra over $\mathbb{Q}$. In the consideration of classes of lattices $I \subset B$, we make a choice and consider lattices as right modules—considerations on the left are analogous, with the map $I \mapsto \overline{I}$ allowing passage between left and right. We say that lattices $I, J \subseteq B$ are **in the same right class**, and write $I \sim_{\mathsf{R}} J$, if there exists $\alpha \in B^\times$ such that $\alpha I = J$; equivalently, $I \sim_{\mathsf{R}} J$ if and only if I is isomorphic to J as right modules over $O_{\mathsf{R}}(I) = O_{\mathsf{R}}(J)$. The relation $\sim_{\mathsf{R}}$ is evidently an equivalence relation, and the class of a lattice I is denoted $[I]_{\mathsf{R}}$.

Let $O \subset B$ be an order. We define the **right class set** of O as

$$\operatorname{Cls}_{\mathsf{R}} O := \{[I]_{\mathsf{R}} : I \subset B \text{ invertible and } O_{\mathsf{R}}(I) = O\};$$

equivalently, $\operatorname{Cls}_{\mathsf{R}} O$ is the set of isomorphism classes of invertible right O-modules in B. The standard involution induces a bijection between $\operatorname{Cls}_{\mathsf{R}} O$ and the analogously defined left class set $\operatorname{Cls}_{\mathsf{L}} O$; working on the right from now on, we will often abbreviate $\operatorname{Cls} O := \operatorname{Cls}_{\mathsf{R}} O$.

© The Author(s) 2021

J. Voight, *Quaternion Algebras*, Graduate Texts in Mathematics 288,
https://doi.org/10.1007/978-3-030-56694-4_17

Unfortunately, the class set $\operatorname{Cls} \mathcal{O}$ does *not* have the structure of a group: only a pointed set, with distinguished element $[\mathcal{O}]_{\mathsf{R}}$. One problem is the compatibility of multiplication discussed in the previous chapter. But even if we allowed products between incompatible lattices, the product need not be well-defined: the lattices IJ and $I\alpha J$ for $\alpha \in B^\times$ need not be in the same class, because of the failure of commutativity. (This is the reason we write 'Cls' instead of 'Cl', as a reminder that it is only a class *set*.) In Chapter 19, we will describe the structure that arises naturally instead: a partially defined product on classes of lattices, a *groupoid*.

In any case, using the same method of proof (geometry of numbers) as in the commutative case, we will show that there exists a constant C (depending on $\mathcal{O}$) such that every class in $\operatorname{Cls} \mathcal{O}$ is represented by an integral ideal $I \subseteq \mathcal{O}$ with $\mathsf{N}(I) = \#(\mathcal{O}/I) \leq C$. As a consequence, we have the following fundamental theorem.

Theorem 17.1.1. *Let B be a quaternion algebra over $\mathbb{Q}$ and let $\mathcal{O} \subset B$ be an order. Then the right class set* $\operatorname{Cls} \mathcal{O}$ *is finite.*

Accordingly, we call $\# \operatorname{Cls} \mathcal{O} \in \mathbb{Z}_{\geq 1}$ the **(right) class number** of $\mathcal{O}$.

Right class sets pass between orders as follows. Let $\mathcal{O}, \mathcal{O}' \subset B$ be orders. If $\mathcal{O} \simeq \mathcal{O}'$ are isomorphic as rings, then of course this isomorphism induces a bijection $\operatorname{Cls} \mathcal{O} \xrightarrow{\sim} \operatorname{Cls} \mathcal{O}'$. In fact, $\mathcal{O} \simeq \mathcal{O}'$ if and only if there exists $\alpha \in B^\times$ such that $\mathcal{O}' = \alpha^{-1}\mathcal{O}\alpha$ by the Skolem–Noether theorem; for historical reasons, we say that $\mathcal{O}, \mathcal{O}'$ are **of the same type**.

Note that $I = \mathcal{O}\alpha = \alpha\mathcal{O}'$ has $\mathcal{O}_{\mathsf{L}}(I) = \mathcal{O}$ and $\mathcal{O}_{\mathsf{R}}(I) = \mathcal{O}'$ (recalling 10.2.5). With this in mind, more generally, we say that $\mathcal{O}'$ is **connected** to $\mathcal{O}$ if there exists an invertible lattice J with $\mathcal{O}_{\mathsf{L}}(J) = \mathcal{O}$ and $\mathcal{O}_{\mathsf{R}}(J) = \mathcal{O}'$, called a **connecting ideal**. Because invertible lattices are locally principal, two orders are connected if and only if they are **locally of the same type** (i.e., **locally isomorphic**). If $\mathcal{O}'$ is connected to $\mathcal{O}$, then right multiplying by a $\mathcal{O}, \mathcal{O}'$-connecting ideal J yields a bijection

$$\begin{aligned} \operatorname{Cls} \mathcal{O} &\xrightarrow{\sim} \operatorname{Cls} \mathcal{O}' \\ [I]_{\mathsf{R}} &\mapsto [IJ]_{\mathsf{R}} \end{aligned} \tag{17.1.2}$$

We define the **genus** of an order $\mathcal{O} \subset B$ to be the set $\operatorname{Gen} \mathcal{O}$ of orders in B locally isomorphic to $\mathcal{O}$, and the **type set** $\operatorname{Typ} \mathcal{O}$ of $\mathcal{O}$ to be the set of R-isomorphism classes of orders in the genus of $\mathcal{O}$. The map

$$\begin{aligned} \operatorname{Cls} \mathcal{O} &\to \operatorname{Typ} \mathcal{O} \\ [I]_{\mathsf{R}} &\mapsto \text{class of } \mathcal{O}_{\mathsf{L}}(I) \end{aligned} \tag{17.1.3}$$

is a surjective map of sets, so the type set is finite: in other words, up to isomorphism, there are only finitely many types of orders in the genus of $\mathcal{O}$. All maximal orders in B are in the same genus, so in particular there are only finitely many conjugacy classes of maximal orders in B. In this way, the right class set of $\mathcal{O}$ also organizes the types of orders arising from $\mathcal{O}$.

The most basic question about the class number is of course its size (as a function of $\mathcal{O}$). In the case of quadratic fields, the behavior of the class group depends in a significant way on whether the field is imaginary or real: for negative discriminant

$d < 0$, the Brauer–Siegel theorem provides that $\#\mathrm{Cl}\, S$ is approximately of size $\sqrt{|d|}$; in contrast, for positive discriminant $d > 0$, one typically sees a small class group and a correspondingly large fundamental unit, but this statement is notoriously difficult to establish unconditionally.

The same dichotomy is at play in the case of quaternion algebras, and to state the cleanest results we suppose that O is a maximal order. Let $D := \mathrm{disc}\, B = \mathrm{discrd}(O)$ be the discriminant of B. If B is definite, which is to say $\infty \in \mathrm{Ram}\, B$, then B is like an imaginary quadratic field K: the norm is positive definite. In this case, $\#\mathrm{Cls}\, O$ is approximately of size D, a consequence of the *Eichler mass formula*, the subject of Chapter 25. On the other hand, if B is indefinite, akin to a real quadratic field, then $\#\mathrm{Cls}\, O = 1$, this time a consequence of *strong approximation*, the subject of Chapter 28. Just as in the commutative case, estimates on the size of the class number use analytic methods and so must wait until we have developed the required tools.

17.2 Matrix ring

To begin, we first consider classes of ideals for the matrix ring; here, we can use methods from linear algebra before we turn to more general methods in the rest of the chapter.

17.2.1. Let R be a PID with field of fractions F, and let $B = \mathrm{M}_n(F)$. By Corollary 10.5.5, every maximal order of $B = \mathrm{M}_n(F)$ is conjugate to $\mathrm{M}_n(R)$. Moreover, every two-sided ideal of $\mathrm{M}_n(R)$ is principal, generated by an element $a \in F^\times$ (multiplying a candidate ideal by matrix units, as in Exercise 7.5(b)), so the group of fractional two-sided $\mathrm{M}_n(R)$-ideals is canonically identified with the group of fractional R-ideals, itself isomorphic to the free abelian group on the (principal) nonzero prime ideals of R.

Just as in the two-sided case, the right class set for $\mathrm{M}_n(R)$ is trivial.

Proposition 17.2.2. *Let R be a PID with field of fractions F, and let $B = \mathrm{M}_n(F)$. Let $I \subseteq B$ be an R-lattice with either $O_{\mathsf{L}}(I)$ or $O_{\mathsf{R}}(I)$ maximal. Then I is principal, and both $O_{\mathsf{L}}(I)$ and $O_{\mathsf{R}}(I)$ are maximal.*

Proof. We may suppose I is integral by rescaling by $r \in R$. Replacing I by the transpose $I^{\mathsf{t}} = \{\alpha^{\mathsf{t}} : \alpha \in I\}$ interchanging left and right orders (Exercise 10.12) if necessary, we may suppose that $O_{\mathsf{L}}(I)$ is maximal. Then, by Corollary 10.5.5, we have $O_{\mathsf{L}}(I) = \alpha^{-1}\,\mathrm{M}_n(R)\alpha$ with $\alpha \in B^\times$, so replacing I by $\alpha^{-1}I$ we may suppose $O_{\mathsf{L}}(I) = \mathrm{M}_n(R)$.

Now we follow Newman [New72, Theorem II.5]. Let $\alpha_1, \dots, \alpha_m$ be R-module generators for I. Consider the $nm \times n$ matrix $A = (\alpha_1, \dots, \alpha_m)^{\mathsf{t}}$. By row reduction over R (Hermite normal form, proven as part of the structure theorem for finitely generated modules over a PID), there exists $Q \in \mathrm{GL}_{nm}(R)$ such that $QA = (\beta, 0)^{\mathsf{t}}$ and $\beta \in \mathrm{M}_n(R)$. We will show that $I = \mathrm{M}_n(R)\beta$. Let $\nu_{11}, \dots, \nu_{1m} \in \mathrm{M}_n(R)$ be the block matrices in the top n rows of Q. Then $\beta = \nu_{11}\alpha_1 + \dots + \nu_{1m}\alpha_m$ so $\beta \in I$ and $\mathrm{M}_n(R)\beta \subseteq I$. Conversely, let $\mu_{11}, \dots, \mu_{m1} \in \mathrm{M}_n(R)$ be the block matrices in the left n columns of $Q^{-1} \in \mathrm{GL}_{nm}(R)$. Since $Q^{-1}(\beta, 0)^{\mathsf{t}} = A$, we have $\mu_{i1}\beta = \alpha_i$ so $\alpha_i \in \mathrm{M}_n(R)\beta$ for

$i = 1, \dots, m$, thus $I \subseteq \mathrm{M}_n(R)\beta$. Therefore $I = \mathrm{M}_n(R)\beta$, and so $\mathcal{O}_{\mathsf{R}}(I)$ is maximal (16.2.3). ◻

Returning to the case of quaternion algebras, we have the following corollary of Proposition 17.2.2.

Corollary 17.2.3. *Let R be a Dedekind domain and let B be a quaternion algebra over $F = \operatorname{Frac} R$. Let $I \subseteq B$ be an R-lattice with either $\mathcal{O}_{\mathsf{L}}(I)$ or $\mathcal{O}_{\mathsf{R}}(I)$ maximal. Then I is locally principal and both $\mathcal{O}_{\mathsf{L}}(I)$ and $\mathcal{O}_{\mathsf{R}}(I)$ are maximal.*

Proof. For each prime $\mathfrak{p}$ of R, we have that $R_{\mathfrak{p}}$ is a DVR and one of two possibilities: either $B_{\mathfrak{p}} \simeq \mathrm{M}_2(F_{\mathfrak{p}})$, in which case we can apply Lemma 17.2.2 to conclude $I_{\mathfrak{p}}$ is principal, or $B_{\mathfrak{p}}$ is a division algebra, and we instead apply 13.3.10 to conclude that $I_{\mathfrak{p}}$ is principal. ◻

17.3 Classes of lattices

For the rest of this chapter, let R be a Dedekind domain with field of fractions $F = \operatorname{Frac} R$, and let B be a simple F-algebra.

Definition 17.3.1. Let $I, J \subseteq B$ be R-lattices. We say I, J are **in the same right class**, and we write $I \sim_{\mathsf{R}} J$, if there exists $\alpha \in B^\times$ such that $\alpha I = J$.

17.3.2. Throughout, we work on the right; analogous definitions can be made on the left. When B has a standard involution, the map $I \mapsto \overline{I}$ interchanges left and right.

Lemma 17.3.3. *Let $I, J \subseteq B$ be R-lattices. Then the following are equivalent:*

(i) *$I \sim_{\mathsf{R}} J$;*
(ii) *I is isomorphic to J as a right module over $\mathcal{O}_{\mathsf{R}}(I) = \mathcal{O}_{\mathsf{R}}(J)$; and*
(iii) *$(J : I)_{\mathsf{L}}$ is a principal R-lattice.*

Proof. For (i) ⇒ (ii). If $I \sim_{\mathsf{R}} J$ then $J = \alpha I$ with $\alpha \in B^\times$, so $\mathcal{O}_{\mathsf{R}}(J) = \mathcal{O}_{\mathsf{R}}(I)$ and the map left-multiplication by α gives a right $\mathcal{O}$-module isomorphism $I \xrightarrow{\sim} J$. Conversely, for (i) ⇐ (ii), suppose that $\phi \colon I \xrightarrow{\sim} J$ is an isomorphism of right $\mathcal{O}$-modules. Then $\phi_F : I \otimes_R F = B \xrightarrow{\sim} J \otimes_R F = B$ is an automorphism of B as a right B-module. Then as in Example 7.2.14, such an isomorphism is obtained by left multiplication by $\alpha \in B^\times$, so by restriction ϕ is given by this map as well.

Next, for (i) ⇒ (iii), suppose $\alpha I = J$ with $\alpha \in B^\times$. Then

$$(J : I)_{\mathsf{L}} = \{\beta \in B : \beta I \subseteq J = \alpha I\} = \alpha \mathcal{O}_{\mathsf{L}}(I)$$

is principal. The converse follows similarly. ◻

The relation $\sim_{\mathsf{R}}$ defines an equivalence relation on the set of R-lattices in B, and the equivalence class of an R-lattice I is denoted $[I]_{\mathsf{R}}$. If I is an invertible R-lattice, then every lattice in the class $[I]_{\mathsf{R}}$ is invertible and we call the class **invertible**.

In view of Lemma 17.3.3(b), we organize classes of lattices by their right orders. Let $\mathcal{O} \subset B$ be an R-order.

Definition 17.3.4. The **(right) class set** of O is

$$\mathrm{Cls}_\mathsf{R}\, O := \{[I]_\mathsf{R} : I \text{ an invertible right fractional } O\text{-ideal}\}.$$

In view of 17.3.2, we will soon abbreviate $\mathrm{Cls}\, O := \mathrm{Cls}_\mathsf{R}\, O$ and drop the subscript $_\mathsf{R}$ from the classes, when no confusion can result.

Remark 17.3.5. The notation $\mathrm{Cl}\, O$ is also used for the class set, but it sometimes means instead the *stably free* class group or some other variant. We use "Cls" to emphasize that we are working with a class *set*.

17.3.6. The set $\mathrm{Cls}_\mathsf{R}\, O$ has a distinguished element $[O]_\mathsf{R} \in \mathrm{Cls}_\mathsf{R}\, O$, so it has the structure of a pointed set (a set equipped with a distinguished element of the set). However, in general it does not have the structure of a group under multiplication: for classes $[I]_\mathsf{R}, [J]_\mathsf{R}$, we have $[\alpha J]_\mathsf{R} = [J]_\mathsf{R}$ for $\alpha \in B^\times$ but we need not have $[I\alpha J]_\mathsf{R} = [IJ]_\mathsf{R}$, because of the lack of commutativity.

17.3.7. An argument similar to the one in Proposition 17.2.2, either arguing locally or with pseudobases (9.3.7), yields the following [CR81, (4.13)].

Let R be a Dedekind domain with $F = \mathrm{Frac}\, R$, and let $I \subseteq B$ be an R-lattice with $O_\mathsf{L}(I) = \mathrm{M}_n(R)$. Then there exists $\beta \in \mathrm{GL}_n(F)$ and fractional ideals $\mathfrak{a}_1, \cdots, \mathfrak{a}_n$ such that

$$I = \mathrm{M}_n(R)\,\mathrm{diag}(\mathfrak{a}_1, \dots, \mathfrak{a}_n)\beta \tag{17.3.8}$$

where $\mathrm{diag}(\mathfrak{a}_1, \dots, \mathfrak{a}_n)$ is the R-module of diagonal matrices with entries in the given fractional ideal. The representation (17.3.8) is called the **Hermite normal form** of the R-module I, because it generalizes the Hermite normal form over a PID (allowing coefficient ideals).

By 9.3.10, the Steinitz class $[\mathfrak{a}_1 \cdots \mathfrak{a}_n] \in \mathrm{Cl}\, R$ is uniquely defined. Switching to the right, this yields a bijection

$$\begin{aligned} \mathrm{Cl}\, R &\xrightarrow{\sim} \mathrm{Cls}_\mathsf{R}(\mathrm{M}_n(R)) \\ [\mathfrak{a}] &\mapsto [\mathrm{diag}(\mathfrak{a}, 1, \dots, 1)\,\mathrm{M}_n(R)]_\mathsf{R} \end{aligned} \tag{17.3.9}$$

17.4 Types of orders

Next, we consider isomorphism classes of orders. Let $O, O' \subseteq B$ be R-orders.

Definition 17.4.1. We say O, O' are **of the same type** if there exists $\alpha \in B^\times$ such that $O' = \alpha^{-1} O \alpha$.

Lemma 17.4.2. *The R-orders O, O' are of the same type if and only if they are isomorphic as R-algebras.*

Proof. If O, O' are of the same type, then they are isomorphic (under conjugation). Conversely, if $\phi\colon O \xrightarrow{\sim} O'$ is an isomorphism of R-algebras, then extending scalars to F we obtain $\phi_F : OF = B \xrightarrow{\sim} B = O'F$ an F-algebra automorphism of B. By the theorem of Skolem–Noether (Corollary 7.7.4), such an automorphism is given by conjugation by $\alpha \in B^\times$, so O, O' are of the same type. □

17.4.3. If $\mathcal{O}, \mathcal{O}'$ are of the same type, then an isomorphism $\mathcal{O} \xrightarrow{\sim} \mathcal{O}'$ induces a bijection $\operatorname{Cls} \mathcal{O} \xrightarrow{\sim} \operatorname{Cls} \mathcal{O}'$ of pointed sets. By Lemma 17.4.2, such an isomorphism is provided by conjugation $\mathcal{O}' = \alpha^{-1}\mathcal{O}\alpha$ for some $\alpha \in B^\times$. The principal lattice $I = \mathcal{O}\alpha = \alpha\mathcal{O}'$ has $\mathcal{O}_\mathsf{L}(I) = \mathcal{O}$ and $\mathcal{O}_\mathsf{R}(I) = \mathcal{O}'$.

Generalizing 17.4.3, the class sets of two orders are in bijection if they are connected, in the following sense.

Definition 17.4.4. $\mathcal{O}$ is **connected** to $\mathcal{O}'$ if there exists a locally principal fractional $\mathcal{O}, \mathcal{O}'$-ideal $J \subseteq B$, called a **connecting ideal**.

The relation of being connected is an equivalence relation on the set of R-orders. If two R-orders $\mathcal{O}, \mathcal{O}'$ are of the same type, then they are connected by a principal connecting ideal (17.4.3).

Definition 17.4.5. We say that $\mathcal{O}, \mathcal{O}'$ are **locally of the same type** or **locally isomorphic** if $\mathcal{O}_\mathfrak{p}$ and $\mathcal{O}'_\mathfrak{p}$ are of the same type (i.e., $\mathcal{O}_\mathfrak{p} \simeq \mathcal{O}'_\mathfrak{p}$) for all primes $\mathfrak{p}$ of R.

Lemma 17.4.6. *The R-orders $\mathcal{O}, \mathcal{O}'$ are connected if and only if $\mathcal{O}, \mathcal{O}'$ are locally isomorphic.*

Proof. Let J be a connecting ideal, a locally principal fractional $\mathcal{O}, \mathcal{O}'$-ideal. Then for all primes $\mathfrak{p}$ of R we have $J_\mathfrak{p} = \mathcal{O}_\mathfrak{p}\alpha_\mathfrak{p}$ with $\alpha_\mathfrak{p} \in B_\mathfrak{p}$, and consequently $\mathcal{O}'_\mathfrak{p} = \mathcal{O}_\mathsf{R}(I_\mathfrak{p}) = \alpha_\mathfrak{p}^{-1}\mathcal{O}_\mathfrak{p}\alpha_\mathfrak{p}$. Therefore $\mathcal{O}$ is locally isomorphic to $\mathcal{O}'$.

Conversely, if $\mathcal{O}, \mathcal{O}'$ are locally isomorphic, then for all primes $\mathfrak{p}$ of R we have $\mathcal{O}'_\mathfrak{p} = \alpha_\mathfrak{p}^{-1}\mathcal{O}_\mathfrak{p}\alpha_\mathfrak{p}$ with $\alpha_\mathfrak{p} \in B_\mathfrak{p}$. Since R is a Dedekind domain, $\mathcal{O}'_\mathfrak{p} = \mathcal{O}_\mathfrak{p}$ for all but finitely many primes $\mathfrak{p}$, so we may take $\alpha_\mathfrak{p} \in \mathcal{O}_\mathfrak{p} = \mathcal{O}'_\mathfrak{p}$ for all but finitely many primes $\mathfrak{p}$. Therefore, there exists an R-lattice I with $I_\mathfrak{p} = \mathcal{O}_\mathfrak{p}\alpha_\mathfrak{p}$ by the local-global principle for lattices, and I is a locally principal fractional $\mathcal{O}, \mathcal{O}'$-ideal. □

Lemma 17.4.7. *If $\mathcal{O}, \mathcal{O}' \subseteq B$ are maximal R-orders, then $\mathcal{O}\mathcal{O}'$ is a $\mathcal{O}, \mathcal{O}'$-connecting ideal.*

The product in Lemma 17.4.7 is not necessarily compatible.

Proof. Since $\mathcal{O}, \mathcal{O}'$ are R-lattices, their product $I := \mathcal{O}\mathcal{O}'$ is an R-lattice. We visibly have $\mathcal{O} \subseteq \mathcal{O}_\mathsf{L}(I)$ and the same on the right; but $\mathcal{O}, \mathcal{O}'$ are maximal, so equality holds and I is a fractional $\mathcal{O}, \mathcal{O}'$-ideal. Finally, I is invertible by Proposition 16.6.15(b), hence locally principal by Main Theorem 16.6.1. □

In analogy with the class set, we make the following definitions.

Definition 17.4.8. Let $\mathcal{O} \subset B$ be an R-order. The **genus** $\operatorname{Gen} \mathcal{O}$ of $\mathcal{O}$ is the set of R-orders in B locally isomorphic to $\mathcal{O}$. The **type set** $\operatorname{Typ} \mathcal{O}$ of $\mathcal{O}$ is the set of isomorphism classes of orders in the genus of $\mathcal{O}$.

17.4.9. The orders in a genus have a common reduced discriminant, since the discriminant can be defined locally and is well-defined on (local) isomorphism classes, by Corollary 15.2.9.

17.4.10. Recalling section 15.5, there is a unique genus of maximal R-orders in a quaternion algebra B—that is to say, every two maximal orders are locally isomorphic—and this genus has a well-defined reduced discriminant equal to $\operatorname{disc}_R B$.

The importance of connected orders is attested to by the following result.

Lemma 17.4.11. *Let $\mathcal{O}, \mathcal{O}'$ be connected R-orders, and let J be a connecting $\mathcal{O}, \mathcal{O}'$-ideal. Then the maps*

$$\begin{aligned} \operatorname{Cls_R} \mathcal{O} &\xrightarrow{\sim} \operatorname{Cls_R} \mathcal{O}' \\ [I]_\mathsf{R} &\mapsto [IJ]_\mathsf{R} \\ [I'J^{-1}]_\mathsf{R} &\leftarrow\!\shortmid [I']_\mathsf{R} \end{aligned}$$

are mutually inverse bijections. In particular, if $\mathcal{O}' \in \operatorname{Gen} \mathcal{O}$ then $\#\operatorname{Cls_R} \mathcal{O} = \#\operatorname{Cls_R} \mathcal{O}'$.

Proof. By definition, J is invertible with $\mathcal{O}_\mathsf{L}(J) = \mathcal{O}$ and $\mathcal{O}_\mathsf{R}(J) = \mathcal{O}'$. Therefore the map $I \mapsto IJ$ induces a bijection between the set of invertible right $\mathcal{O}$-ideals and the set of invertible right $\mathcal{O}'$-ideals (Lemma 16.5.11), with inverse given by $I' \mapsto I'J^{-1}$, and each of these products is compatible. This map then induces a bijection $\operatorname{Cls} \mathcal{O} \xrightarrow{\sim} \operatorname{Cls} \mathcal{O}'$, since is compatible with left multiplication in B, i.e., $(\alpha I)J = \alpha(IJ)$ for all $\alpha \in B^\times$. ◻

Remark 17.4.12. The equivalence in Lemma 17.4.11 is a form of *Morita equivalence*: see Remark 7.2.20.

Lemma 17.4.11 says that the cardinality of the right class set is well-defined on the genus $\operatorname{Gen} \mathcal{O}$; and of course the cardinality of the type set is also well-defined on the genus (as it is the number of isomorphism classes).

Lemma 17.4.13. *The map*

$$\begin{aligned} \operatorname{Cls_R} \mathcal{O} &\to \operatorname{Typ} \mathcal{O} \\ [I]_\mathsf{R} &\mapsto \text{class of } \mathcal{O}_\mathsf{L}(I) \end{aligned} \tag{17.4.14}$$

is a surjective map of sets.

Proof. If $\mathcal{O}'$ is connected to $\mathcal{O}$, then there is a connecting $\mathcal{O}', \mathcal{O}$-ideal I, and $[I]_\mathsf{R} \in \operatorname{Cls_R} \mathcal{O}$ has $\mathcal{O}_\mathsf{L}(I) \simeq \mathcal{O}'$. ◻

Remark 17.4.15. The fibers of the map (17.4.14) is given by classes of two-sided ideals: see Proposition 18.5.10.

17.4.16. Let $B = \mathrm{M}_2(F)$ and $\mathcal{O} = \mathrm{M}_2(R)$. From the bijection (17.3.9), the classes in $\operatorname{Cl_R}(\mathrm{M}_2(R))$ are represented by $I_\mathfrak{a} = \begin{pmatrix} \mathfrak{a} & \mathfrak{a} \\ R & R \end{pmatrix}$ for $[\mathfrak{a}] \in \operatorname{Cl} R$. Consequently

$$\mathcal{O}_\mathsf{L}(I_\mathfrak{a}) = \begin{pmatrix} R & \mathfrak{a} \\ \mathfrak{a}^{-1} & R \end{pmatrix}.$$

We will see later (28.5.11) that there is a bijection

$$\begin{aligned} \operatorname{Cl} R/(\operatorname{Cl} R)^2 &\xrightarrow{\sim} \operatorname{Typ} \mathrm{M}_2(R) \\ \text{class of } [\mathfrak{a}] \text{ up to squares} &\mapsto \text{class of } \begin{pmatrix} R & \mathfrak{a} \\ \mathfrak{a}^{-1} & R \end{pmatrix}. \end{aligned} \tag{17.4.17}$$

17.5 ▹ Finiteness of the class set: over the integers

Over the next two sections, we will show that the set $\operatorname{Cls} O$ of invertible right (fractional) O-ideals is finite using the geometry of numbers. In this section, we carry this out for the simplest case, when B is definite over $\mathbb{Q}$; we consider the general case in the next section. For further reading on the rich theory of the geometry of numbers, see Cassels [Cas97], Gruber–Lekkerkerker [GrLe87], and Siegel [Sie89].

Our strategy is as follows: if J is an invertible right O-ideal, we will show there exists $\alpha \in J^{-1}$ with the property that $\alpha J = I \subseteq O$ has bounded absolute norm $\mathsf{N}(I) = \#(O/I) \leq C$ where $C \in \mathbb{R}_{>0}$ is independent of J. The result will then follow from the fact that there are only finitely many right O-ideals of bounded absolute norm.

We begin with some definitions (generalizing Definition 9.3.1 slightly).

Definition 17.5.1. A **Euclidean lattice** is a $\mathbb{Z}$-submodule $\Lambda \subseteq \mathbb{R}^n$ with $\Lambda \simeq \mathbb{Z}^n$ such that $\mathbb{R}\Lambda = \mathbb{R}^n$. The **covolume** of a Euclidean lattice Λ is $\operatorname{covol}(\Lambda) = \operatorname{vol}(\mathbb{R}^n/\Lambda)$.

17.5.2. Equivalently, a Euclidean lattice $\Lambda \subset \mathbb{R}^n$ is the $\mathbb{Z}$-span of a basis of $\mathbb{R}^n$, and if $\Lambda = \bigoplus_i \mathbb{Z}a_i$, then $\operatorname{covol}(\Lambda) = |\det(a_{ij})_{i,j}|$.

Lemma 17.5.3. *A subgroup $\Lambda \subset \mathbb{R}^n$ is a Euclidean lattice if and only if Λ is discrete and the quotient $\mathbb{R}^n/\Lambda$ is compact.*

Proof. Exercise 17.6. □

Definition 17.5.4. Let $X \subseteq \mathbb{R}^n$ be a subset.

(a) X is **convex** if $tx + (1-t)y \in X$ for all $x, y \in X$ and $t \in [0, 1]$.

(b) X is **symmetric** if $-x \in X$ for all $x \in X$.

The main result of Minkowski's geometry of numbers is the following convex body theorem.

Theorem 17.5.5. (Minkowski). *Let $X \subseteq \mathbb{R}^n$ be a closed, convex, symmetric subset of $\mathbb{R}^n$, and let $\Lambda \subset \mathbb{R}^n$ be a Euclidean lattice. If $\operatorname{vol}(X) \geq 2^n \operatorname{covol}(\Lambda)$, then there exists $0 \neq \alpha \in \Lambda \cap X$.*

The following proposition can be seen as a generalization of what was done for the Hurwitz order (11.3.1).

Proposition 17.5.6. *Let B be a definite quaternion algebra over $\mathbb{Q}$ and let $O \subset B$ be an order. Then $O^\times = O^1$ is a finite group, and every right ideal class in* $\operatorname{Cls} O$ *is represented by an integral right O-ideal with*

$$\mathsf{N}(I) \le \frac{8}{\pi^2}\operatorname{discrd}(O)$$

and the right class set $\operatorname{Cls} O$ *is finite.*

Proof. Let $B = \left(\dfrac{a,b}{\mathbb{Q}}\right)$, with $a, b \in \mathbb{Z}_{<0}$. Since B is definite, there is an embedding $B \hookrightarrow B_\infty = B \otimes_{\mathbb{Q}} \mathbb{R} \simeq \mathbb{H}$. Inside $B_\infty \simeq \mathbb{R}^4$ with Euclidean norm nrd, the order O sits as a Euclidean lattice. The set O^1 is therefore a discrete subset of the compact set $B_\infty^1 \simeq \mathbb{H}^1$, so it is finite.

Explicitly, we identify

$$\begin{aligned} B_\infty &\xrightarrow{\sim} \mathbb{R}^4 \\ t + xi + yj + zij &\mapsto \sqrt{2}\big(t, x\sqrt{|a|}, y\sqrt{|b|}, z\sqrt{|ab|}\big) \end{aligned} \tag{17.5.7}$$

Then $2\operatorname{nrd}(\alpha) = \|\alpha\|^2$ for $\alpha \in B$ in this identification, and we have $\operatorname{covol}(O) = \operatorname{discrd}(O)$ (Exercise 17.7).

Let $J \subset B$ be an invertible right fractional O-ideal. To find I with $[I] = [J]$ and I integral, we look for a small $\alpha \in J^{-1}$ so that $I = \alpha J \subseteq O$ will do. As a measure of (co)volume, counting cosets and applying the definition (16.4.9), we obtain

$$\operatorname{covol}(J^{-1}) = [O : J^{-1}]_{\mathbb{Z}} \operatorname{covol}(O) = \mathsf{N}(J^{-1}) \operatorname{discrd}(O). \tag{17.5.8}$$

Let $c > 0$ satisfy $c^4 = (32/\pi^2)\operatorname{covol}(J^{-1})$, and let

$$X = \{x \in \mathbb{R}^4 : \|x\| \le c\}.$$

Then X is closed, convex, and symmetric, and $\operatorname{vol}(X) = \pi^2 c^4/2 = 16\operatorname{covol}(J^{-1})$. Then by Minkowski's theorem (Theorem 17.5.5), there exists $0 \ne \alpha \in J^{-1} \cap X$, and

$$\begin{aligned} \mathsf{N}(\alpha J) = \operatorname{Nm}_{B|\mathbb{Q}}(\alpha)\mathsf{N}(J) = \operatorname{nrd}(\alpha)^2\mathsf{N}(J) &= \frac{1}{4}\|\alpha\|^4\mathsf{N}(J) \\ \le \frac{1}{4}c^4\mathsf{N}(J) = \frac{8}{\pi^2}\operatorname{discrd}(O). \end{aligned} \tag{17.5.9}$$

Since α is nonzero and B is a division algebra, $\alpha \in B^\times$. Since $\alpha \in J^{-1}$, the integral right fractional O-ideal $I = \alpha J \subseteq O$ is as desired.

If $I \subseteq O$ has $\mathsf{N}(I) = \#(O/I) \le C$ for $C \in \mathbb{Z}_{>0}$, then $CO \subseteq I \subseteq O$ hence there are only finitely many possibilities for I, and the second statement follows. □

17.6 ▹ Example

We pause for an extended example. We steal the following lemma from the future.

Lemma 17.6.1. *Let $e \in \mathbb{Z}_{\geq 0}$. Then every principal right* $\mathrm{M}_2(\mathbb{Z}_p)$*-ideal I with* $\mathrm{nrd}(I) = p^e$ *is of the form $I = \alpha\,\mathrm{M}_2(\mathbb{Z}_p)$ where*

$$\alpha \in \left\{ \begin{pmatrix} p^u & 0 \\ c & p^v \end{pmatrix} : u, v \in \mathbb{Z}_{\geq 0}, u + v = e, and c \in \mathbb{Z}/p^v\mathbb{Z} \right\}. \tag{17.6.2}$$

Proof. The lemma follows from the theory of invariant factors: a more general statement is proven in Lemma 26.4.1. ◻

Example 17.6.3. Let $B = \left(\dfrac{-1, -23}{\mathbb{Q}}\right)$, and let

$$O = \mathbb{Z} + \mathbb{Z}i + \mathbb{Z}\frac{1+j}{2} + \mathbb{Z}i\frac{1+j}{2}.$$

We have $\mathrm{discrd}(O) = \mathrm{disc}\, B = 23$, so O is a maximal order, and $\beta = (1+j)/2$ satisfies $\beta^2 - \beta + 6 = 0$. For convenience, let $\alpha = i$, so $O = \mathbb{Z}\langle \alpha, \beta \rangle$. Then

$$\alpha\beta + \beta\alpha = \alpha. \tag{17.6.4}$$

By Proposition 17.5.6, it is sufficient to compute the (invertible) right O-ideals $I \subseteq O$ such that

$$\mathrm{nrd}(I)^2 = \mathsf{N}(I) \leq \frac{8}{\pi^2}(23) \leq 18.7$$

so $\mathrm{nrd}(I) \leq 4$. For $\mathrm{nrd}(I) = 1$, we can only have $I = O$, and the class $[I_1] = [O]$. Let $O_1 = O$.

We move to $\mathrm{nrd}(I) = 2$, and refer to Lemma 17.6.1. Since B is split at 2, there is an embedding

$$\begin{aligned} O &\hookrightarrow \mathrm{M}_2(\mathbb{Z}_2) \\ \alpha, \beta &\mapsto \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}, \begin{pmatrix} 1 & 0 \\ 0 & b_0 \end{pmatrix}. \end{aligned}$$

where $b_0 = 2 + 8 + 16 + 32 + \cdots \in \mathbb{Z}_2$ satisfies $b_0^2 - b_0 + 6 = 0$ and $b_0 \equiv 0 \pmod{2}$. We have

$$\beta, \beta + 1, (\alpha + 1)\beta \equiv \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}, \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix}, \begin{pmatrix} 1 & 0 \\ 1 & 0 \end{pmatrix} \pmod{2}$$

so we obtain the three right ideals

$$I_{(1:0)} = 2O + \beta O, \quad I_{(0:1)} = 2O + (\beta - 1)O, \quad I_{(1:1)} = 2O + (\alpha + 1)\beta O \tag{17.6.5}$$

labelled by the corresponding nonzero column. If one of these three ideals is principal, then it is generated by an element of reduced norm 2. We have

$$\begin{aligned} &\mathrm{nrd}(t + x\alpha + y\beta + z\alpha\beta) \\ &\quad = t^2 + ty + x^2 + xz + 6y^2 + 6z^2 \\ &\quad = \left(t + \frac{1}{2}y\right)^2 + \left(x + \frac{1}{2}z\right)^2 + \frac{23}{4}y^2 + \frac{23}{4}z^2. \end{aligned} \tag{17.6.6}$$

So $\operatorname{nrd}(\gamma) = 2$ with $\gamma \in \mathcal{O}$ has $t, x, y, z \in \mathbb{Z}$ and therefore $y = z = 0$ and $t = x = 1$, i.e., $I_{(1:1)} = (\alpha + 1)\mathcal{O}$ is principal, and the ideals $I_{(1:0)}, I_{(0:1)}$ are not. But $[I_{(1:0)}] = [I_{(0:1)}]$ because $\alpha I_{(1:0)} = I_{(0:1)}$ because $\alpha \in \mathcal{O}^\times$ and by (17.6.4)

$$\alpha(2\mathcal{O} + (\beta - 1)\mathcal{O}) = 2\alpha\mathcal{O} + \alpha(\beta - 1)\mathcal{O} = 2\mathcal{O} - \beta\alpha\mathcal{O} = I_{(0:1)}$$

(We have $\alpha I_{(1:0)} \neq I_{(1:0)}$ precisely because $\alpha \notin \mathcal{O}_\mathsf{L}(I)$.) In this way, we have found exactly one new right ideal class, $[I_2] = [I_{(1:0)}]$. We compute its left order to be

$$\mathcal{O}_2 := \mathcal{O}_\mathsf{L}(I_2) = \mathbb{Z} + \beta\mathbb{Z} + \frac{i(1+3j)}{4}\mathbb{Z} + (2ij)\mathbb{Z} \neq \mathcal{O}$$

and we also have a new type $[\mathcal{O}_2] \neq [\mathcal{O}_1] \in \operatorname{Typ} \mathcal{O}$.

In a similar way, we find 4 right ideals of reduced norm 3, and exactly one new right ideal class, represented by the right ideal $I_3 = 3\mathcal{O} + (\alpha + 1)\beta\mathcal{O}$. For example, we find that the right ideal $I' = 3\mathcal{O} + \beta\mathcal{O}$ is not principal using (17.6.6): letting

$$(I' : I_2)_\mathsf{L} = I' I_2^{-1} = \frac{1}{2} I' \overline{I_2}$$

and we find a shortest vector

$$(1 - \beta)/2 \in (I' : I_2)_\mathsf{L},$$

so $[I'] = [I_2]$.

Repeating this with ideals of reduced norm 4 (Exercise 17.8), we conclude that

$$\operatorname{Cls} \mathcal{O} = \{[I_1], [I_2], [I_3]\}$$

and letting $\mathcal{O}_3 := \mathcal{O}_\mathsf{L}(I_3)$, checking it is not isomorphic to the previous two orders, we have

$$\operatorname{Typ} \mathcal{O} = \{[\mathcal{O}_1], [\mathcal{O}_2], [\mathcal{O}_3]\}.$$

17.7 Finiteness of the class set: over number rings

We now turn to the general case.

Main Theorem 17.7.1. *Let F be a number field, let $S \subseteq \operatorname{Pl} F$ be eligible and $R = R_{(S)}$ be the ring of S-integers in F. Let B be a quaternion algebra over F, and let $\mathcal{O} \subseteq B$ be an R-order in B. Then the class set* $\operatorname{Cls} \mathcal{O}$ *and the type set* $\operatorname{Typ} \mathcal{O}$ *is finite.*

We call $\# \operatorname{Cls} \mathcal{O}$ the **(right) class number** of $\mathcal{O}$. (By 17.3.2, the left class number suitably defined is equal to the right class number.) This result will be drastically improved upon in Part III of this text from analytic considerations; the proof in this section, using the geometry of numbers, has the advantage that is easy to visualize, it works in quite some generality, and it is the launching point for algorithmic aspects.

17.7.2. Before we begin, two quick reductions. The finiteness of the type set follows from finiteness of the right class set by Lemma 17.4.13. And if $R = \mathbb{Z}_F$ is the ring of integers of F, then the general case follows from the fact that the map

$$\begin{aligned} \operatorname{Cls} O &\to \operatorname{Cl}(O \otimes_R R_{(S)}) \\ [I] &\mapsto [I \otimes_R R_{(S)}] \end{aligned} \tag{17.7.3}$$

is surjective for an eligible set S.

Let F be a number field of degree $n = [F : \mathbb{Q}]$, let $R = \mathbb{Z}_F$ be the ring of integers in F, and let B be a quaternion algebra over F.

17.7.4. Suppose that F has r real places and c complex places, so that $n = r + 2c$. Then

$$F \hookrightarrow F_\infty = F \otimes_{\mathbb{Q}} \mathbb{R} \simeq \prod_{v \mid \infty} F_v \simeq \mathbb{R}^r \times \mathbb{C}^c. \tag{17.7.5}$$

Taking the basis $1, i$ for $\mathbb{C}$, we obtain $F_\infty \simeq \mathbb{R}^n$, and then in the embedding (17.7.5), the ring of integers $R \simeq \mathbb{Z}^n$ sits discretely inside $F_\infty \simeq \mathbb{R}^n$ as a Euclidean lattice.

17.7.6. Suppose $B = \left(\dfrac{a,b}{F}\right)$ and let $1, i, j, k$ be the standard basis for B with $k = ij$, so $B = F \oplus Fi \oplus Fj \oplus Fk \simeq F^4$ as F-vector spaces. Then

$$B \hookrightarrow B_\infty = B \otimes_{\mathbb{Q}} \mathbb{R} \simeq B \otimes_F F_\infty \simeq F_\infty^4 \tag{17.7.7}$$

in this same basis. Via (17.7.5) in each of the four components, the embedding (17.7.7) then gives an identification $B_\infty \simeq (\mathbb{R}^n)^4 \simeq \mathbb{R}^{4n}$.

The order $R\langle i, j, k\rangle = R + Ri + Rj + Rk$ is discrete in B_∞ exactly because R is discrete in F. But then implies that an R-order O is discrete in B_∞, since $[O : R\langle i, j, k\rangle]_{\mathbb{Z}} < \infty$. Therefore $O \hookrightarrow \mathbb{R}^{4n}$ has the structure of a Euclidean lattice.

In the previous section, the real vector space B_∞ was Euclidean under the reduced norm. In general, that need no longer be the case. Instead, we find a positive definite quadratic form $Q\colon B_\infty \to \mathbb{R}$ that **majorizes** the reduced norm in the following sense: we require that

$$|\operatorname{Nm}_{F/\mathbb{Q}}(\operatorname{nrd}(\alpha))| \le Q(\alpha)^n \tag{17.7.8}$$

for all $\alpha \in B \subseteq B_\infty$.

Remark 17.7.9. With respect to possible majorants (17.7.8): in general, there are uncountably many such choices, and parametrizing majorants arises in a geometric context as part of *reduction theory*. As it will turn out, the only "interesting" case to consider here is 17.7.10, by strong approximation (see Theorem 17.8.3).

17.7.10. Let B be a totally definite (Definition 14.5.7) quaternion algebra over F, a totally real number field. Then the quadratic form

$$\begin{aligned} Q\colon B &\to \mathbb{Q} \\ \alpha &\mapsto \operatorname{Tr}_{F/\mathbb{Q}}(\operatorname{nrd}(\alpha)) = \sum_{v \mid \infty} v(\operatorname{nrd}(\alpha)) \end{aligned} \tag{17.7.11}$$

is positive definite: $B_v \simeq \mathbb{H}$ and so $v(\operatorname{nrd}(\alpha)) \geq 0$ with equality if and only if $\alpha = 0$. We call this quadratic form the **absolute reduced norm**. In this case, by the arithmetic-geometric mean,

$$\begin{aligned} \operatorname{Nm}_{F/\mathbb{Q}}(\operatorname{nrd}(\alpha))^{1/n} &= \left(\prod_v v(\operatorname{nrd}(\alpha))\right)^{1/n} \\ &\leq \frac{1}{n}\sum_v v(\operatorname{nrd}(\alpha)) = \frac{1}{n}Q(\alpha) \end{aligned} \tag{17.7.12}$$

(with equality if and only if $v(\operatorname{nrd}\alpha)$ agrees for all v).

We pause to note the following important consequence of 17.7.10.

Lemma 17.7.13. *Let B be a totally definite quaternion algebra over a totally real field F and let $O \subseteq B$ be a $\mathbb{Z}_F$-order. Then the group of units of reduced norm* 1

$$O^1 = \{\gamma \in O : \operatorname{nrd}(\gamma) = 1\}$$

is a finite group.

In Lemma 17.7.13, if $F = \mathbb{Q}$ then $O^\times = O^1$, so we have captured the entire unit group.

Proof. As in 17.7.10, we equip $B_\mathbb{R} := B \otimes_\mathbb{Q} \mathbb{R} \simeq \mathbb{H}^n \simeq \mathbb{R}^{4n}$ with the absolute reduced norm giving $O \hookrightarrow B_\mathbb{R}$ the structure of a Euclidean lattice (17.7.7). We have

$$O^1 = \{\gamma \in O : Q(\gamma) = n\} \tag{17.7.14}$$

by the arithmetic-geometric mean (17.7.12). But the set $\{x \in B_\mathbb{R} : Q(x) = n\}$ is an ellipsoid in $\mathbb{R}^{4n}$ so compact, and O is a lattice so discrete. Therefore the intersection O^1 is finite. ◻

17.7.15. We now generalize 17.7.10 to the general case. For v an infinite place of F, define

$$\begin{aligned} Q_v \colon B_v &\to \mathbb{R} \\ t + xi + yj + zij &\mapsto |v(t)|^2 + |v(a)||v(x)|^2 + |v(b)||v(y)|^2 + |v(ab)||v(z)|^2; \end{aligned}$$

then Q_v is a positive definite quadratic form on B_v, and

$$\begin{aligned} |v(\operatorname{nrd}(\alpha))| &= |v(t^2 - ax^2 - by^2 + abz^2)| \\ &\leq |v(t)|^2 + |v(a)||v(x)|^2 + |v(b)||v(y)|^2 + |v(ab)||v(z)|^2 \\ &= Q_v(\alpha). \end{aligned} \tag{17.7.16}$$

Let $m_v = 1, 2$ depending on if v is real or complex, and define

$$\begin{aligned} Q \colon B_\infty \simeq \prod_{v\mid\infty} B_v &\to \mathbb{R} \\ (\alpha_v)_v &\mapsto \sum_{v\mid\infty} m_v Q_v(\alpha_v). \end{aligned} \tag{17.7.17}$$

Then Q is a positive definite quadratic form on B_∞, again called the **absolute reduced norm** (relative to a, b); it depends on the choice of representation $B = \left(\dfrac{a,b}{F}\right)$. Nevertheless, (17.7.16) and the arithmetic-geometric mean yield

$$\begin{aligned} |\mathrm{Nm}_{F/\mathbb{Q}}(\mathrm{nrd}(\alpha))|^{1/n} &\le \frac{1}{n}\sum_{v\mid\infty} m_v |v(\mathrm{nrd}(\alpha))| \\ &\le \frac{1}{n}\sum_{v\mid\infty} m_v Q_v(\alpha) = Q(\alpha). \end{aligned} \tag{17.7.18}$$

We are now ready to prove the main result of this section.

Proposition 17.7.19. *There exists an explicit constant $C \in \mathbb{R}_{>0}$ such that for all R-orders O, every right ideal class in* $\mathrm{Cls}\, O$ *is represented by an integral right O-ideal I with*

$$\mathsf{N}(I) \le C\mathsf{N}(\mathrm{discrd}(O)).$$

Proof. If $B \simeq \mathrm{M}_2(F)$, then we appeal to 17.3.7, where such a bound comes from the finiteness of $\mathrm{Cl}\, R$. So we may suppose that B is a division ring.

Let

$$X = \{(x_i)_i \in \mathbb{R}^{4n} : Q(\alpha) \le 1\}. \tag{17.7.20}$$

Then X is closed, convex, and symmetric.

Let O be an R-order in B and let J be an invertible right fractional O-ideal. As in (17.5.8), counting cosets gives

$$\mathrm{covol}(J^{-1}) = \mathsf{N}(J)^{-1}\,\mathrm{covol}(O). \tag{17.7.21}$$

Let

$$c := 2\left(\frac{\mathrm{covol}(J^{-1})}{\mathrm{vol}(X)}\right)^{1/4n}. \tag{17.7.22}$$

Then $\mathrm{vol}(cX) = c^{4n}\,\mathrm{vol}(X) = 2^{4n}\,\mathrm{covol}(J^{-1})$. By Minkowski's theorem, there exists $0 \ne \alpha \in J^{-1} \cap cX$, so $Q(\alpha) \le c^2$. By (17.7.18),

$$|\mathrm{Nm}_{F/\mathbb{Q}}(\mathrm{nrd}(\alpha))| \le \frac{1}{n^n} Q(\alpha)^n \le \frac{c^{2n}}{n^n}.$$

Consequently

$$\begin{aligned} \mathsf{N}(\alpha J) = |\mathrm{Nm}_{F/\mathbb{Q}}(\mathrm{nrd}(\alpha))|^2 \mathsf{N}(J) &\le \frac{c^{4n}}{n^{2n}}\mathsf{N}(J) \\ &= \frac{2^{4n}\mathsf{N}(J)^{-1}\,\mathrm{covol}(O)}{n^{2n}\,\mathrm{vol}(X)}\mathsf{N}(J) = \frac{2^{4n}\,\mathrm{covol}(O)}{n^{2n}\,\mathrm{vol}(X)} \\ &= C\mathsf{N}(\mathrm{discrd}(O)) \end{aligned} \tag{17.7.23}$$

with

$$C := \frac{2^{4n}}{n^{2n}\,\mathrm{vol}(X)} \frac{\mathrm{covol}(O)}{\mathsf{N}(\mathrm{discrd}(O))}. \tag{17.7.24}$$

The ratio $\operatorname{covol}(O)/\mathrm{N}(\operatorname{discrd}(O))$ is a constant independent of O: if O' is another R-order then

$$\frac{\mathrm{N}(\operatorname{discrd}(O'))}{\operatorname{covol}(O')} = \frac{[O : O']_{\mathbb{Z}}\mathrm{N}(\operatorname{discrd}(O))}{[O : O']_{\mathbb{Z}} \operatorname{covol}(O)} = \frac{\mathrm{N}(\operatorname{discrd}(O))}{\operatorname{covol}(O)}.$$

Since α is nonzero and B is a division algebra we conclude that $\alpha \in B^{\times}$, and since $\alpha \in J^{-1}$, the ideal $I = \alpha J$ is as desired. □

Remark 17.7.25. For an explicit version of the Minkowski bound in the totally definite case, with a careful choice of compact region, see Kirschmer [Kir2005, Theorem 3.3.11].

Lemma 17.7.26. *For all $C > 0$, there are only finitely many integral right O-ideals with $\mathrm{N}(I) \leq C$.*

Proof. We may suppose $C \in \mathbb{Z}$. If $I \subseteq O$ then $\mathrm{N}(I) = [O : I]_{\mathbb{Z}} \leq C$, so $CO \subseteq I \subseteq O$. But the group O/CO is a finite abelian group and there are only finitely many possibilities for I. □

We now have the ingredients for our main theorem.

Proof of Main Theorem 17.7.1. Combine Proposition 17.7.19, the reductions in 17.7.2, and Lemma 17.7.26. □

Remark 17.7.27. The finiteness statement (Main Theorem 17.7.1) can be generalized to the following theorem of Jordan–Zassenhaus. Let R be a Dedekind domain with $F = \operatorname{Frac}(R)$ a global field, let $O \subseteq B$ be an R-order in a finite-dimensional semisimple algebra B, and let V be a left B-module. Then there are only finitely many isomorphism classes $I \subseteq B$ with $O \subseteq O_{\mathsf{L}}(I)$. Specializing to $V = B$ a quaternion algebra, we recover the Main Theorem 17.7.1. For a proof, see Reiner [Rei2003, Theorem 26.4]; see also the discussion by Curtis–Reiner [CR81, §24].

17.8 Eichler's theorem

In this section, we state a special but conceptually important case of Eichler's theorem for number fields: roughly speaking, the class set of an indefinite quaternion order is in bijection with a certain class group of the base ring.

Let F be a number field with ring of integers $R = \mathbb{Z}_F$ and let B be a quaternion algebra over F.

Definition 17.8.1. We say B **satisfies the Eichler condition** if B is indefinite.

Definition 17.8.1 introduces a longer (and rather opaque) phrase for something that we already had a word for, but its use is prevalent in the literature. There are two options: either B is totally definite (F is a totally real field and all archimedean places of F are ramified in B) or B is indefinite and satisfies the Eichler condition.

17.8.2. Recall 14.7.2 that we define $\Omega \subseteq \operatorname{Ram} B$ to be the set of real ramified places of B and $F^\times_{>_\Omega 0}$ to be the positive elements for $v \in \Omega$.

We now define the group $\operatorname{Cl}_\Omega R$ as

the group of fractional ideals of F under multiplication

modulo

the subgroup of nonzero principal fractional ideals generated by an element in $F^\times_{>_\Omega 0}$

If Ω is the set of all real places of F, then $\operatorname{Cl}_\Omega R = \operatorname{Cl}^+ R$ is the **narrow** (or **strict**) **class group**. On the other hand, if $\Omega = \emptyset$, then $\operatorname{Cl}_\Omega R = \operatorname{Cl} R$. In general, we have surjective group homomorphisms $\operatorname{Cl}^+ R \to \operatorname{Cl}_\Omega R$ and $\operatorname{Cl}_\Omega R \to \operatorname{Cl} R$. In the language of class field theory, $\operatorname{Cl}_\Omega R$ is the class group corresponding to the cycle given by the product of the places in Ω.

Theorem 17.8.3. (Eichler; strong approximation). *Let F be a number field and let B be a quaternion algebra over F that satisfies the Eichler condition. Let $O \subseteq B$ be a maximal $\mathbb{Z}_F$-order. Then the reduced norm induces a bijection*

$$\begin{aligned} \operatorname{Cls} O &\xrightarrow{\sim} \operatorname{Cl}_\Omega R \\ [I] &\mapsto [\operatorname{nrd}(I)]. \end{aligned} \tag{17.8.4}$$

where $\Omega \subseteq \operatorname{Ram} B$ is the set of real ramified places in B.

Proof. Eichler's theorem is addressed by Reiner [Rei2003, §34], with a global proof of the key result [Rei2003, Theorem 34.9] falling over several pages. We will instead prove a more general version of this theorem as part of strong approximation, when idelic methods allow for a more efficient argument: see Corollary 28.5.17. □

Eichler's theorem says that when B is *not* totally definite, the only obstruction for an ideal to be principal in a maximal order is that its reduced norm fails to be (strictly) principal in the base ring. In particular, we have the following corollary.

Corollary 17.8.5. *If $\#\operatorname{Cl}^+ R = 1$, then $\#\operatorname{Cls} O = 1$: i.e., every right O-ideal of a maximal order in an indefinite quaternion algebra is principal.*

Proof. Immediate from Eichler's theorem and the fact that $\operatorname{Cl}^+ R$ surjects onto $\operatorname{Cl}_\Omega R$, by 17.8.2. □

Corollary 17.8.6. *There is a bijection* $\operatorname{Cls} \mathrm{M}_2(\mathbb{Z}_F) \xrightarrow{\sim} \operatorname{Cl} \mathbb{Z}_F$.

Proof. Immediate from Eichler's theorem; we proved this more generally for a matrix ring (17.3.9) using the Hermite normal form. □

17.8.7. It is sensible for the class group $\operatorname{Cl}_\Omega R$ to appear by norm considerations. Let $v \in \Omega$; then $B_v \simeq \mathbb{H}$, and if $\alpha \in B^\times$ then $v(\operatorname{nrd}(\alpha)) > 0$, as the reduced norm is positive.

The class sets of totally definite orders are not captured by Eichler's theorem, and for good reason: they can be arbitrarily large, a consequence of the Eichler mass formula (Chapter 25).

Exercises

Unless otherwise specified, throughout these exercises let R be a Dedekind domain with field of fractions F, let B be a quaternion algebra over F, and let $O \subseteq B$ be an R-order.

1. Argue for Proposition 17.2.2 directly in a special case as follows. Let $I \subseteq \mathrm{M}_2(F)$ be a lattice with $O_{\mathsf{R}}(I) = \mathrm{M}_2(R)$.

   (a) By considering $I \otimes_R F$ show that
   $$I \subseteq \begin{pmatrix} F & F \\ 0 & 0 \end{pmatrix} \mathrm{M}_2(R) \oplus \begin{pmatrix} 0 & 0 \\ F & F \end{pmatrix} \mathrm{M}_2(R).$$

   (b) Suppose that R is a PID. Conclude that I is principal.

2. Let $O, O' \subseteq B$ be R-orders. Show that the map in Lemma 17.4.11 is a bijection of *pointed* sets if and only if O is isomorphic to O'.

▹3. Let $O, O' \subseteq B$ be R-orders with $O \subseteq O'$.

   (a) If I is an invertible right O-ideal, show that IO' is an invertible right O'-ideal. (The product IO' is not necessarily compatible.)

   (b) Show that the map
   $$\begin{aligned} \operatorname{Cls} O &\to \operatorname{Cls} O' \\ [I] &\mapsto [IO'] \end{aligned}$$
   is well-defined, surjective, and has finite fibers. *[Hint: let $r \in R$ be nonzero such that $O' \subseteq r^{-1}O$. If $IO' = I'$, then $I' = IO' \subseteq r^{-1}I \subseteq r^{-1}I'$ so $rI' \subseteq I \subseteq I'$, and conclude there are only finitely many possibilities for I.]*

4 Let $O, O' \subseteq B$ be maximal R-orders. In this exercise, we prove the following statement:

   > There is a unique *integral* connecting O, O' ideal I of minimal reduced norm; moreover, we have $\operatorname{nrd}(I) = [O : O \cap O']$.

   (a) Show that this statement is local, i.e., the statement is true over R if and only if it is true over $R_{\mathfrak{p}}$ for all primes $\mathfrak{p}$ of R.
   (b) Suppose R is a DVR. Show that the statement is true if B is a division algebra.
   (c) Suppose R is a DVR with maximal ideal $\mathfrak{p}$, and that $B \simeq \mathrm{M}_2(F)$. Show that there is a unique $\alpha \in O \smallsetminus \mathfrak{p}O$ such that $O' = \alpha^{-1}O\alpha$ up to left multiplication by $O^\times$, and conclude that $I = O\alpha$ is the unique integral connecting O, O' ideal of minimal reduced norm. *[Hint: $N_{\mathrm{GL}_2(F)}(\mathrm{M}_2(R)) = F^\times \mathrm{GL}_2(R)$.]*
   (d) Continuing (c), show that $\operatorname{nrd}(\alpha) = [O : O \cap O']$. *[Hint: the statement is equivalent under left or right multiplication of α* by $O^\times \simeq \mathrm{GL}_2(R)$, *so consider invariant factors.]* [For another perspective, see section 23.5.]

▹5. Let $\mathcal{O} \subseteq B$ be an R-order and let I be an invertible fractional right $\mathcal{O}$-ideal. Let $\mathfrak{a} \subseteq R$ be a nonzero ideal. Show that there exists a representative $J \in [I]_{\mathsf{R}}$ (in the same right ideal class as I) such that $J \subseteq \mathcal{O}$ and $\operatorname{nrd}(J)$ is coprime to $\mathfrak{a}$. *[Hint: look for $\alpha \in (\mathcal{O} : I)_{\mathsf{R}}$ and then look locally.]*

▹6. Prove Lemma 17.5.3: a subgroup $\Lambda \subset \mathbb{R}^n$ is a Euclidean lattice if and only if Λ is discrete (every point of Λ is isolated, i.e., every $x \in \Lambda$ has an open neighborhood $U \ni x$ such that $\Lambda \cap U = \{x\}$) and the quotient $\mathbb{R}^n/\Lambda$ is compact.

▹7. Let B be a definite quaternion algebra over $\mathbb{Q}$ and let $\mathcal{O} \subset B$ be an order.

(a) Let $B_\infty = B \otimes_{\mathbb{Q}} \mathbb{R}$. Show that nrd is a Euclidean norm on B_∞, and $\mathcal{O}$ is discrete in B_∞ with $\operatorname{covol}(\mathcal{O}) = 4\operatorname{discrd}(\mathcal{O})$. [So it is better to take $\sqrt{2}\operatorname{nrd}$ instead, to get $\operatorname{covol}(\mathcal{O}) = \operatorname{discrd}(\mathcal{O})$ on the nose.]
(b) Let $K_1, K_2 \subseteq B$ be quadratic fields contained in B with $K_1 \cap K_2 = \mathbb{Q}$. Let $S_i := K_i \cap \mathcal{O}$ and $d_i = \operatorname{disc} S_i$. Show that

$$(|d_1| - 1)(|d_2| - 1) \geq 4\operatorname{discrd}(\mathcal{O}).$$

[Hint: write $S_i = \mathbb{Z}[\alpha_i]$ and consider the order $\mathbb{Z}\langle \alpha_1, \alpha_2\rangle$.]
(c) Prove that if $\alpha_1, \alpha_2 \in \mathcal{O}$ have

$$\operatorname{nrd}(\alpha_1), \operatorname{nrd}(\alpha_2) < \frac{\sqrt{\operatorname{discrd}(\mathcal{O})}}{2}$$

then $\alpha_1\alpha_2 = \alpha_2\alpha_1$.

8. Complete Example 17.6.3 by showing explicitly that all right $\mathcal{O}$-ideals of reduced norm 4 are in the same right ideal class as one of I_1, I_2, I_3.

9. Let R be a global ring with $\#\operatorname{Cl} R = 1$, i.e., every fractional R-ideal is principal $\mathfrak{a} = aR$. Suppose further that $\#\operatorname{Cls}\mathcal{O} = 1$. Let $\alpha \in \mathcal{O}$ have $\operatorname{nrd}(\alpha) \neq 0$, and factor $\operatorname{nrd}(\alpha) = \pi_1\pi_2\cdots\pi_r \in R$ where $\pi_i \in R$ are pairwise nonassociate nonzero prime elements (equivalently $\pi_i R$ are pairwise distinct nonzero prime ideals).

(a) Show that there exist $\varpi_1, \varpi_2, \dots, \varpi_r \in \mathcal{O}$ such that $\alpha = \varpi_1\varpi_2\cdots\varpi_r$ and $\operatorname{nrd}(\varpi_i)R = \pi_i R$ for all $i = 1, \dots, r$.

(b) Show that every other such factorization is of the form

$$\alpha = (\varpi_1\gamma_1)(\gamma_1^{-1}\varpi_2\gamma_2)\cdots(\gamma_{r-1}^{-1}\varpi_r)$$

where $\gamma_1, \dots, \gamma_r \in \mathcal{O}^\times$.

(c) Suppose that $\operatorname{nrd}(\mathcal{O}^\times) = R^\times$. Refine part (a) and show that the stronger conclusion that there exist ϖ_i such that $\operatorname{nrd}(\varpi_i) = \pi_i$ for all i.

[This generalizes Theorem 11.4.8.]

10. We have seen that maximal orders in (definite) quaternion algebras of discriminant 2 (the Hurwitz order) and discriminant 3 (Exercise 11.11) are Euclidean with respect to the norm, and in particular they have trivial right class set.

(a) Show that maximal orders O in quaternion algebras of discriminants 5, 7, 13 have $\#\operatorname{Cls} O = 1$.
(b) Conclude that the quaternary quadratic forms

$$t^2 + tx + ty + tz + x^2 + xy + xz + 2y^2 - yz + 2z^2,$$
$$t^2 + tz + x^2 + xy + 2y^2 + 2z^2,$$
$$t^2 + ty + tz + 2x^2 + xy + 2xz + 2y^2 + yz + 4z^2$$

are multiplicative and **universal**, i.e., represent all positive integers.
(c) Show that for discriminant 7, 13 the maximal orders are *not* Euclidean with respect to the norm.

[We discuss the maximal orders of class number 1 in Theorem 25.4.1. The maximal order for discriminant 5 is in fact norm Euclidean: see Fitzgerald [Fit2011].]

11. In this exercise, we show that the group of principal two-sided ideals $\operatorname{PIdl}(O)$ need not be normal in the group of invertible fractional O-ideals $\operatorname{Idl}(O)$ of an order.

Let $B = (-1, -1 \mid \mathbb{Q})$, and let $O \subseteq B$ be the Hurwitz order. Let $O' = \mathbb{Z} + 5O = O(5)$ (cf. Exercise 18.6). Show that

$$I' = 10O' + (1 - 2i + j)O'$$

is a two-sided invertible O'-ideal, and that

$$I' j (I')^{-1} = 5O' + (i + 3j + k)O'$$

is not principal.

12. The finiteness of the class group (see Reiner [Rei2003, Lemma 26.3]) can be proven replacing the geometry of numbers with just the pigeonhole principle, as follows. Let B be a division algebra over a number field F with ring of integers R, and let $O \subseteq B$ be an R-order.

(a) To prove the finiteness of $\operatorname{Cls} O$, show that without loss of generality we may take $F = \mathbb{Q}$.
(b) Show that $\operatorname{Nm}_{B|\mathbb{Q}}(x_1\alpha_1 + \cdots + x_n\alpha_n) \in \mathbb{Q}[x_1, \ldots, x_n]$ is a homogeneous polynomial of degree n.
(c) Show that there exists $C \in \mathbb{Z}_{>0}$ such that for all $t > 0$ and all $x \in \mathbb{Z}^n$ with $|x_i| \leq t$, we have $|\operatorname{Nm}_{B|F}(x_1\alpha_1 + \cdots + x_n\alpha_n)| \leq Ct^n$.
(d) Let $I \subseteq O$ be a lattice. Let $s \in \mathbb{Z}$ be such that

$$s^n \leq \mathsf{N}(I) = \#(O/I) \leq (s+1)^n.$$

Using the pigeonhole principle, show that there exists $\alpha = \sum_i x_i\alpha_i \in I$ with $x_i \in \mathbb{Z}$ and $|x_i| \leq 2(s+1)$ for all i.
(e) Show that $\mathsf{N}(\alpha O) \leq 2^n(s+1)^n C$, and conclude that

$$\#(I/\alpha O) \leq 4^n C.$$

(f) Let $M = (4^n C)!$ and show that $MI \subseteq \alpha O$, whence

$$MO \subseteq I' \subseteq O$$

where $I' = (M\alpha^{-1})I$. Conclude that the number of possibilities for I' is finite, hence the number of right classes of lattices $I \subseteq O$ is finite, and hence $\# \operatorname{Cls} O < \infty$.

This chapter is licensed under the terms of the Creative Commons Attribution-NonCommercial 4.0 International License (http://creativecommons.org/licenses/by-nc/4.0/), which permits any noncommercial use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license and indicate if changes were made.

The images or other third party material in this chapter are included in the chapter's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the chapter's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.

Chapter 18

Two-sided ideals and the Picard group

In this chapter, we treat maximal orders like noncommutative Dedekind domains, and we consider the structure of *two-sided* ideals (and their classes), in a manner parallel to the commutative case.

18.1 ▹ Noncommutative Dedekind domains

Let R be a Dedekind domain with field of fractions F: then by definition R is noetherian, integrally closed, and all nonzero prime ideals of R are maximal. Equivalently, every ideal of R is the product of prime ideals (uniquely up to permutation). To establish this latter property of unique factorization of ideals, there are two essential ingredients: first, every proper ideal contains a finite product of prime ideals, and second, every nonzero prime ideal $\mathfrak{p} \subseteq R$ is invertible. The first of these uses that R is noetherian and that nonzero prime ideals of R are maximal; the second uses that R is integrally closed.

Here, the theorems are no easier to prove in the case of a quaternion algebra, so we might as well consider them in more generality. Let B be a simple F-algebra and let $O \subseteq B$ be an R-order.

To draw the closest analogy with Dedekind domains, we suppose that $O \subset B$ is *maximal*: this is the noncommutative replacement for *integrally closed*. Since O is finitely generated, if $I \subseteq O$ is a two-sided O-ideal, then I is a finitely generated R-submodule, so the noetherian condition on R automatically implies that every chain of ideals of O stabilizes. We say a two-sided ideal $P \subseteq O$ is **prime** if $P \neq O$ and for all two-sided ideals $I, J \subseteq O$, we have

$$IJ \subseteq P \quad \Rightarrow \quad I \subseteq P \text{ or } J \subseteq P.$$

Running parallel to the above, we have the following initial lemma.

Lemma 18.1.1. *A nonzero two-sided O-ideal is prime if and only it is maximal, and every two-sided O-ideal contains a product of prime two-sided O-ideals.*

Completing the analogy with the commutative case, we then have the following theorem.

© The Author(s) 2021

J. Voight, *Quaternion Algebras*, Graduate Texts in Mathematics 288,
https://doi.org/10.1007/978-3-030-56694-4_18

Theorem 18.1.2. *Let R be a Dedekind domain with field of fractions $F = \operatorname{Frac} R$, let B be a simple F-algebra and let $O \subseteq B$ be a maximal R-order. Then the following statements hold.*

(a) *If $I \subseteq B$ is an R-lattice such that $O_{\mathsf{L}}(I) = O$ or $O_{\mathsf{R}}(I) = O$, then I is invertible and both $O_{\mathsf{L}}(I)$ and $O_{\mathsf{R}}(I)$ are maximal R-orders.*

(b) *Multiplication of two-sided O-ideals is commutative, and every nonzero two-sided O-ideal is the product of finitely many prime two-sided O-ideals, uniquely up to permutation.*

Let $\operatorname{Idl}(O)$ be the group of invertible two-sided fractional O-ideals. Put another way, Theorem 18.1.2 says that if O is maximal, then $\operatorname{Idl}(O)$ is isomorphic to the free abelian group on the set of nonzero prime two-sided O-ideals under multiplication.

We now consider classes of two-sided ideals, in the spirit of section 17.1. Two candidates present themselves. On the one hand, inside the group $\operatorname{Idl}(O)$ of invertible fractional two-sided O-ideals, the principal fractional two-sided O-ideals (those of the form $O\alpha O = O\alpha = \alpha O$ for certain $\alpha \in B^\times$) form a subgroup $\operatorname{PIdl}(O)$, and we could consider the quotient. On the other hand, for a commutative ring S, the **Picard group** $\operatorname{Pic}(S)$ is defined to be the group of isomorphism classes of rank one projective (equivalently, invertible) S-modules under the tensor product. When S is a Dedekind domain, there is a canonical isomorphism $\operatorname{Cl} S \cong \operatorname{Pic}(S)$.

For simplicity, suppose now that $R = \mathbb{Z}$. In this noncommutative setting, we analogously define $\operatorname{Pic} O$ to be the group of isomorphism classes of invertible O-bimodules (over $\mathbb{Z}$) under tensor product. If $I, J \in \operatorname{Idl}(O)$, then I, J are isomorphic as O-bimodules if and only if $J = aI$ with $a \in \mathbb{Q}^\times$, and this yields an isomorphism

$$\operatorname{Pic} O \simeq \operatorname{Idl}(O)/\mathbb{Q}^\times.$$

Let

$$N_{B^\times}(O) = \{\alpha \in B^\times : \alpha O = O\alpha\}$$

be the normalizer of O in B. By the Skolem–Noether theorem,

$$N_{B^\times}(O)/\mathbb{Q}^\times \simeq \operatorname{Aut}(O)$$

is the group of $\mathbb{Z}$-algebra (or ring) automorphisms of O.

Theorem 18.1.3. *Let B be a quaternion algebra over $\mathbb{Q}$ of discriminant $D := \operatorname{disc} B$, and let $O \subset B$ be a maximal order. Then*

$$\operatorname{Pic} O \simeq \prod_{p \mid D} \mathbb{Z}/2\mathbb{Z}$$

generated by (unique) prime two-sided O-ideals with reduced norm $p \mid D$, and there is an exact sequence

$$\begin{gathered} 0 \to N_{B^\times}(O)/(\mathbb{Q}^\times O^\times) \to \operatorname{Pic} O \to \operatorname{Idl}(O)/\operatorname{PIdl}(O) \to 0 \\ \alpha(\mathbb{Q}^\times O^\times) \mapsto [O\alpha O]. \end{gathered} \tag{18.1.4}$$

In particular, Pic O is a finite abelian 2-group. As an application of Theorem 18.1.3, we revisit the map (17.1.3):

$$\begin{aligned} \operatorname{Cls} O &\to \operatorname{Typ} O \\ [I]_{\mathsf{R}} &\mapsto \text{class of } O_{\mathsf{L}}(I) \end{aligned}$$

We recall that this map is surjective. The fibers are given by Theorem 18.1.3 (see Proposition 18.5.10): the fiber above the isomorphism class of O' is in bijection with the set $\operatorname{PIdl}(O') \backslash \operatorname{Idl}(O')$.

Remark 18.1.5. The structure of Pic O is more complicated when O is not necessarily a maximal order: in general, the group Pic O is finite but it may be nonabelian (see Exercise 18.6); worse still, in general the subgroup $\operatorname{PIdl}(O)$ may not be a normal subgroup in $\operatorname{Idl}(O)$.

18.2 Prime ideals

Throughout this chapter, let R be a Dedekind domain with field of fractions $F = \operatorname{Frac} R$, let B be a simple finite-dimensional F-algebra, and let $O \subseteq B$ be an R-order.

18.2.1. Let $I \subseteq O$ be a nonzero two-sided ideal. In view of Remark 16.2.10, we see that I is automatically an R-lattice: $IF \subseteq B$ is a two-sided ideal of B, so since B is simple and $I \neq \{0\}$ we must have $IF = B$.

Definition 18.2.2. A two-sided ideal $P \subseteq O$ is **prime** if $P \neq O$ and for all two-sided ideals $I, J \subseteq O$ we have

$$IJ \subseteq P \quad \Rightarrow \quad I \subseteq P \text{ or } J \subseteq P.$$

A two-sided O-ideal $M \subseteq O$ is **maximal** if $M \neq O$ and M is not properly contained in another two-sided ideal.

Example 18.2.3. The zero ideal $P = \{0\}$ is prime: see Exercise 18.2.

18.2.4. Let $P \subseteq O$ be a two-sided ideal. Then the two-sided O/P-ideals are in bijection with the two-sided O-ideals containing P. If $P \neq O$, then P is prime if and only if for all two sided O/P-ideals $I/P, J/P$, we have

$$(I/P)(J/P) = \{0\} \quad \Rightarrow \quad I/P = \{0\} \text{ or } J/P = \{0\}. \tag{18.2.5}$$

Lemma 18.2.6. *If M is a maximal two-sided O-ideal, then M is prime.*

Proof. Suppose $IJ \subseteq M$. Then $(I + M)(J + M) \subseteq M$. But $I + M \supseteq M$ so either $I + M = M$ or $I + M = O$ by maximality, and the same is true for J. Since $M \neq O$ we must have either $I + M = M$ or $J + M = M$, which is to say $I \subseteq M$ or $J \subseteq M$. □

Proposition 18.2.7.

(a) *A nonzero two-sided O-ideal is prime if and only it is maximal.*

(b) *If $P \subseteq O$ is a nonzero prime two-sided O-ideal, then $\mathfrak{p} = P \cap R$ is a nonzero prime ideal of R, and O/P is a finite-dimensional simple algebra over the field $R/\mathfrak{p}$.*

Proof. We follow Reiner [Rei2003, Theorem 22.3]. The implication ($\Rightarrow$) in (a) follows from Lemma 18.2.6. Conversely, let P be a nonzero prime two-sided O-ideal, and let $\mathfrak{p} = P \cap R$. We show $\mathfrak{p}$ is a nonzero prime. By 18.2.1, P is an R-lattice, so $\mathfrak{p} \neq \{0\}$; since $1 \notin P$, we have $\mathfrak{p} \neq R$ and $\mathfrak{p}$ is nontrivial. If $a, b \in R$, then $ab \in \mathfrak{p}$ implies $(aO)(bO) \subseteq P$; since P is prime, we have $aO \subseteq P$ or $bO \subseteq P$, so $a \in \mathfrak{p}$ or $b \in \mathfrak{p}$.

Now let $J/P = \operatorname{rad}(O/P)$ be the Jacobson radical of O/P (see section 7.4). By Lemma 7.4.8, the ideal J/P is nilpotent; by (18.2.5), we conclude $J/P = \{0\}$. Thus O/P is semisimple by Lemma 7.4.2 and thus is a product of simple $R/\mathfrak{p}$-algebras by the Wedderburn–Artin theorem (Main Theorem 7.3.10). But the simple components of O/P are two-sided ideals that annihilate one another; again by (18.2.5), there can be only one component, and O/P is simple. Thus O/P has no nontrivial ideals, and P is maximal. ∎

Lemma 18.2.8. *Every nonzero two-sided ideal of O contains a (finite) product of nonzero prime ideals.*

Proof. If not, then the set of ideals which do not contain such products is nonempty; since O is noetherian, there is a maximal element M. Since M cannot itself be prime, there exist ideals I, J, properly containing M, such that $IJ \subseteq M$. But both I, J contain products of prime ideals, so the same is true of M, a contradiction. ∎

We now turn to notions of invertibility.

18.2.9. Let I be an invertible two-sided fractional O-ideal (cf. Definition 16.2.9 and 16.5.17). In particular, $O_{\mathsf{L}}(I) = O_{\mathsf{R}}(I) = O$. If J is another invertible two-sided fractional O-ideal, then so is IJ, by Lemma 16.5.11: we have $O_{\mathsf{L}}(IJ) = O_{\mathsf{L}}(I) = O$ and $O_{\mathsf{R}}(IJ) = O_{\mathsf{R}}(J) = O$. Let $\operatorname{Idl}(O)$ be the set of invertible two-sided fractional O-ideals. Then $\operatorname{Idl}(O)$ is a group under multiplication with identity element O.

The structure of $\operatorname{Idl}(O)$, and quotients under natural equivalence relations, is the subject of this chapter.

18.3 Invertibility

We now consider invertibility first in the general context of orders, then for maximal orders. The general theory of maximal orders over Dedekind domains in simple algebras was laid out by Auslander–Goldman [AG60]. One of the highlights of this theory are the classification of such orders: they are endomorphism rings of a finitely generated projective module over a maximal order in a division algebra. For a quite general treatment of maximal orders, see the book by Reiner [Rei2003]; in particular, the ideal theory presented here is also discussed in Reiner [Rei2003, §§22–23].

Lemma 18.3.1. *Let J be a two-sided O-ideal, not necessarily invertible. If $J \subsetneq O$, then $J^{-1} \supsetneq O$.*

Proof. The R-lattice J^{-1} has $J^{-1} \supseteq O$ and $O_{\mathsf{L}}(J^{-1}) \subseteq O_{\mathsf{R}}(J) = O$ and the same result holds interchanging left and right.

We follow Reiner [Rei2003, Lemma 23.4] (who calls the proof "mystifying"). Assume for the purposes of contradiction that $J^{-1} = O$. Since $J \subsetneq O$, there exists a maximal two-sided O-ideal $M \supseteq J$. Thus $M^{-1} \subseteq J^{-1} = O$. By Lemma 18.2.6, M is prime. Let $a \in R \cap J^{-1}$ be nonzero. By Lemma 18.2.8, aO contains a product of prime two-sided O-ideals, so

$$M \supseteq aO \supseteq P_1 P_2 \cdots P_r,$$

with each P_i prime. We may suppose without loss of generality that $r \in \mathbb{Z}_{>0}$ is minimal with this property. Since $P_1 \cdots P_r \subseteq M$ and M is prime, we must have $P_i \subseteq M$, so $P_i = M$ by Proposition 18.2.7. Thus

$$M \supseteq aO \supseteq J_1 M J_2$$

with J_1, J_2 two-sided O-ideals. From $a^{-1} J_1 M J_2 \subseteq O$, we have $J_1(a^{-1} M J_2) J_1 \subseteq J_1$, so by definition $a^{-1} M J_2 J_1 \subseteq O_{\mathsf{L}}(J_1) = O$. Thus $M(a^{-1} J_2 J_1) M \subseteq M$ and $a^{-1} J_2 J_1 \subseteq M^{-1} \subseteq O$ so $J_2 J_1 \subseteq aO$. This shows that aO contains the product $J_2 J_1$ of $r - 1$ prime two-sided O-ideals, contradicting the minimality of r. □

Using this lemma, we arrive at the following proposition for maximal orders.

Proposition 18.3.2. *Let $I \subseteq B$ be an R-lattice such that $O_{\mathsf{L}}(I)$ is a maximal R-order. Then I is right invertible, i.e., $II^{-1} = O_{\mathsf{L}}(I)$.*

Of course, one can also swap left for right in the statement of Proposition 18.3.2. Using the standard involution, we proved Proposition 18.3.2 when B is a quaternion algebra (Proposition 16.6.15(b)).

Proof of Proposition 18.3.2. We follow Reiner [Rei2003, Theorem 23.5]. Let $O = O_{\mathsf{L}}(I)$. Let $J = II^{-1} \subseteq O$. Then $JI = II^{-1}I \subseteq I$, so $J \subseteq O_{\mathsf{L}}(I) = O$ and J is a two-sided O-ideal. We have

$$JJ^{-1} = II^{-1}J^{-1} \subseteq O,$$

so $I^{-1}J^{-1} \subseteq I^{-1}$ and therefore $J^{-1} \subseteq O_{\mathsf{R}}(I^{-1})$. Additionally,

$$O_{\mathsf{R}}(I^{-1}) \supseteq O_{\mathsf{L}}(I) = O; \tag{18.3.3}$$

but O is maximal, so equality holds in (18.3.3) and therefore $J^{-1} \subseteq O$. But $O \subseteq J^{-1}$ as well, so $J^{-1} = O$. If $J \subsetneq O$, then we have a contradiction with Lemma 18.3.1; so $J = O$, and the proof is complete. □

Putting these ingredients together, we have the following theorem.

Theorem 18.3.4. *Let R be a Dedekind domain with $F = \operatorname{Frac} R$, let B be a simple F-algebra, and let $O \subseteq B$ be a maximal R-order. Then:*

(a) *Multiplication of two-sided ideals is commutative: if I, J are two-sided O-ideals, then $IJ = JI$.*
(b) *Every nonzero two-sided O-ideal is invertible and uniquely expressible as a product of prime two-sided ideals in O.*

Proof. For (b), invertibility follows from Proposition 18.3.2. For (b) without uniqueness, assume for purposes of contradiction that there is a two-sided ideal of O that is not the product of prime ideals; then there is a maximal counterexample J. Since J is not prime, there exists a prime Q with $J \subsetneq Q \subsetneq O$, so $J \subset JQ^{-1} \subsetneq O$. If $J = JQ^{-1}$, so by cancelling $Q = O$, a contradiction. Therefore $JQ^{-1} = P_1 \cdots P_r$ is the product of primes by maximality, and $J = P_1 \cdots P_r Q$ is the product of primes, a contradiction.

We now prove (a). If $P, Q \subseteq O$ are distinct nonzero prime two-sided ideals, and we let $Q' = P^{-1}QP$, then $Q' \subseteq P^{-1}OP = O$ is prime and $PQ' = QP \subseteq Q$, so $P \subseteq Q$ or $Q' \subseteq Q$; but equality would hold in each case by maximality, and since $P \neq Q$, we must have $Q' = Q$, and multiplication is commutative.

Finally, uniqueness of the factorization in (b) follows as in the commutative case. If $P_1 \cdots P_r = Q_1 \cdots Q_s$, then $P_1 = Q_i$ for some i; multiplying by P_1^{-1} and repeating the argument, we find that $\{P_1, \dots, P_r\} = \{Q_1, \dots, Q_s\}$, and the result follows. □

Corollary 18.3.5. *With hypotheses as in Theorem 18.3.4, the group* $\mathrm{Idl}(O)$ *is isomorphic to the free abelian group on the set of nonzero prime ideals.*

With these arguments in hand, we have the following foundational result for quaternion orders.

Theorem 18.3.6. *Suppose that R is a Dedekind domain. Let B be a quaternion algebra over F and let $O \subseteq B$ be a maximal R-order. Then the map*

$$\begin{aligned} \{\text{Prime two-sided } O\text{-ideals}\} &\leftrightarrow \{\text{Prime ideals of } R\} \\ P &\mapsto P \cap R \end{aligned} \tag{18.3.7}$$

is a bijection.

Moreover, if R is a global ring, then there is an exact sequence

$$\begin{aligned} 0 \to \mathrm{Idl}(R) \to \mathrm{Idl}(O) \to \prod_{\mathfrak{p} \mid \mathfrak{D}} \mathbb{Z}/2\mathbb{Z} \to 0 \\ \mathfrak{a} \mapsto O\mathfrak{a}O \qquad\qquad \end{aligned} \tag{18.3.8}$$

where $\mathfrak{D} = \mathrm{disc}_R(B)$.

Proof. The map (18.3.7) is defined by Proposition 18.2.7, and it is surjective because $\mathfrak{p}O \subseteq P$ is contained in a maximal therefore prime ideal.

Next we show that the map is injective. Let P be a prime ideal, and work with completions at a prime $\mathfrak{p}$. Then $P_\mathfrak{p} = P \otimes_R R_\mathfrak{p} \subseteq O_\mathfrak{p}$ is a maximal ideal of $O_\mathfrak{p}$. If $B_\mathfrak{p} \simeq \mathrm{M}_2(F_\mathfrak{p})$, so $O_\mathfrak{p} \simeq \mathrm{M}_2(R_\mathfrak{p})$, then the only maximal two-sided ideal is $\mathfrak{p}O_\mathfrak{p}$; if instead $B_\mathfrak{p}$ is a division algebra, then there is a unique maximal two-sided ideal $P_\mathfrak{p}$ with $P_\mathfrak{p}^2 = \mathfrak{p}O_\mathfrak{p}$ by Theorem 13.3.11. We can also describe this uniformly, by the proof of Proposition 18.2.7: in all cases, we have $P_\mathfrak{p} = \mathrm{rad}(O_\mathfrak{p})$.

There is a natural group homomorphism

$$\begin{aligned} \mathrm{Idl}(R) &\to \mathrm{Idl}(\mathcal{O}) \\ \mathfrak{a} &\mapsto \mathcal{O}\mathfrak{a}\mathcal{O} = \mathfrak{a}\mathcal{O} \end{aligned}$$

This map is injective, since if $\mathfrak{a}\mathcal{O} = \mathcal{O}$ then $\mathfrak{a}^2 = \mathrm{nrd}(\mathfrak{a}\mathcal{O}) = \mathrm{nrd}(\mathcal{O}) = R$, so $\mathfrak{a} = R$. The cokernel of the map is determined by the previous paragraph. □

Remark 18.3.9. Many of the theorems stated in this section (and chapter) hold more generally for *hereditary orders*: this notion is pursued in Chapter 21. To see what this looks like in a more general context, see Curtis–Reiner [CR81, §26B]. A very general context in which one can make an argument like in section 18.3 was axiomatized by Asano; for an exposition and several references, see McConnell–Robson [McCR87, Chapter 5].

18.4 Picard group

We now proceed to consider classes of two-sided ideals. We begin with a natural but abstract definition, in terms of bimodules. (Recall 20.3.7, that a bimodule is *over* R if the R-action on left and right are equal.)

Definition 18.4.1. The **Picard group** of $\mathcal{O}$ over R is the group $\mathrm{Pic}_R(\mathcal{O})$ of isomorphism classes of invertible $\mathcal{O}$-bimodules over R under tensor product.

Remark 18.4.2. Some authors also write $\mathrm{Picent}(\mathcal{O}) = \mathrm{Pic}_{Z(\mathcal{O})}(\mathcal{O})$ when considering the Picard group over the center of $\mathcal{O}$, the most important case. To avoid additional complication, in this section we suppose that B is *central* over F, so $\mathrm{Pic}_R(\mathcal{O}) = \mathrm{Picent}(\mathcal{O})$.

18.4.3. If $I \subseteq B$ is an R-lattice that is a fractional two-sided $\mathcal{O}$-ideal, then I is a $\mathcal{O}$-bimodule over R. Conversely, if I is a $\mathcal{O}$-bimodule over R then $I \otimes_R F \simeq B$ as B-bimodules, and choosing such an isomorphism gives an embedding $I \hookrightarrow B$ as an R-lattice.

Lemma 18.4.4. *Let $I, J \subseteq B$ be R-lattices that are fractional two-sided $\mathcal{O}$-ideals. Then I is isomorphic to J as $\mathcal{O}$-bimodules over R if and only if there exists $a \in F^\times$ such that $J = aI$.*

Proof. See Exercise 18.9. □

18.4.5. By 18.4.3, there is a natural surjective map

$$\mathrm{Idl}(\mathcal{O}) \to \mathrm{Pic}_R(\mathcal{O});$$

we claim that the kernel of this map is $\mathrm{PIdl}(R) \trianglelefteq \mathrm{Idl}(\mathcal{O})$. By Lemma 19.5.1, every isomorphism class of invertible $\mathcal{O}$-bimodule is represented by an invertible R-lattice $I \subseteq B$, unique up to scaling by $F^\times$, and if $a \in F^\times$ then $a\mathcal{O} = \mathcal{O}$ if and only if

$a \in R \cap \mathcal{O}^\times = R^\times$, so the ideal $aR \in \mathrm{PIdl}(R)$ is well-defined. Thus, we obtain a natural isomorphism

$$\mathrm{Idl}(\mathcal{O})/\mathrm{PIdl}(R) \xrightarrow{\sim} \mathrm{Pic}_R(\mathcal{O}). \tag{18.4.6}$$

Equivalently, the sequence

$$1 \to R^\times \to F^\times \to \mathrm{Idl}(\mathcal{O}) \to \mathrm{Pic}_R(\mathcal{O}) \to 1$$

is exact. One might profitably take (18.4.6) as the definition of $\mathrm{Pic}_R(\mathcal{O})$.

18.4.7. If $\mathcal{O}'$ is locally isomorphic $\mathcal{O}$ (so they are in the same genus), then there is a $\mathcal{O}, \mathcal{O}'$-connecting ideal J, and the map

$$\begin{aligned} \mathrm{Idl}(\mathcal{O}) &\to \mathrm{Idl}(\mathcal{O}') \\ I &\mapsto J^{-1}IJ \end{aligned}$$

is an isomorphism of groups restricting to the identity on $\mathrm{Cl}\, R$, so from (18.4.6) we obtain an isomorphism

$$\mathrm{Pic}_R(\mathcal{O}) \simeq \mathrm{Pic}_R(\mathcal{O}'),$$

analogous to Lemma 17.4.11.

Our remaining task in this section is to examine the structure of $\mathrm{Pic}_R(\mathcal{O})$, and to this end we suppose that B is a quaternion algebra over F.

18.4.8. Suppose that $\mathcal{O}$ is a maximal R-order with $F = \mathrm{Frac}\, R$ a global field. Then taking the quotient by $\mathrm{PIdl}(R)$ in the first two terms in (18.3.8) yields an exact sequence

$$0 \to \mathrm{Cl}\, R \to \mathrm{Pic}_R(\mathcal{O}) \to \prod_{\mathfrak{p} \mid \mathfrak{D}} \mathbb{Z}/2\mathbb{Z} \to 0. \tag{18.4.9}$$

Although this sequence need not split, it does show that the Picard group of the maximal order $\mathcal{O}$ is not far from the class group $\mathrm{Cl}\, R$, the difference precisely measured by the primes that ramify in B.

In general, for a quaternion R-order $\mathcal{O}$ we have the following result.

Proposition 18.4.10. $\mathrm{Pic}_R(\mathcal{O})$ *is a finite group.*

Proof. If $\mathcal{O}$ is maximal, we combine (18.4.9) with the finiteness of $\mathrm{Cl}\, R$ and the fact that there are only finitely many primes $\mathfrak{p}$ dividing the discriminant $\mathfrak{D}$.

Now let $\mathcal{O}$ be an R-order. Then there exists a maximal R-order $\mathcal{O}' \supseteq \mathcal{O}$. We argue as in Exercise 17.3. We define a map of sets:

$$\begin{aligned} \mathrm{Pic}_R(\mathcal{O}) &\to \mathrm{Pic}_R(\mathcal{O}') \\ [I] &\mapsto [\mathcal{O}'I\mathcal{O}'] \end{aligned}$$

The class up to scaling by $F^\times$ is well-defined, and $I' := \mathcal{O}'I\mathcal{O}' \supseteq I$ an R-lattice with left and right orders containing $\mathcal{O}'$, but since $\mathcal{O}'$ is maximal these orders equal $\mathcal{O}'$ and I' is invertible.

By the first paragraph, by finiteness of $\mathrm{Pic}_R(O')$, after rescaling we may suppose I' is one of finitely many possibilities. But there exists nonzero $r \in R$ such that $rO' \subset O$, so

$$I' = O'IO' \subseteq (r^{-1}O)I(r^{-1}O) = r^{-2}I \subseteq r^{-2}I'$$

so $r^2I' \subseteq I \subseteq I'$; since I'/r^2I' is a finite group, this leaves only finitely many possibilities for I. □

Remark 18.4.11. The study of the Picard group is quite general. It was studied in detail by Fröhlich [Frö73]; see also Curtis–Reiner [CR87, §55].

18.5 Classes of two-sided ideals

In this section, we compare the Picard group to the group of "ideals modulo principal ideals".

Let $\mathrm{PIdl}(O) \leq \mathrm{Idl}(O)$ be the subgroup of principal two-sided fractional O-ideals (invertible by 16.5.4). Let

$$N_{B^\times}(O) = \{\alpha \in B^\times : \alpha^{-1}O\alpha = O\}$$

be the normalizer of O in $B^\times$.

Lemma 18.5.1. *There is an exact sequence of groups*

$$\begin{aligned} 1 \to O^\times \to N_{B^\times}(O) &\to \mathrm{PIdl}(O) \to 1 \\ \alpha &\mapsto O\alpha O. \end{aligned} \tag{18.5.2}$$

Proof. We have $\alpha \in N_{B^\times}(O)$ if and only if $\alpha O = O\alpha$ if and only if $O\alpha O$ is a principal two-sided fractional O-ideal, as in Exercise 16.17; this gives a surjective group homomorphism $N_{B^\times}(O) \to \mathrm{PIdl}(O)$. The kernel is the set of $\alpha \in B^\times$ such that $\alpha O = O$, and this normal subgroup is precisely $O^\times$. □

Proposition 18.5.3. *There is an isomorphism of groups*

$$\begin{aligned} N_{B^\times}(O)/(F^\times O^\times) &\xrightarrow{\sim} \mathrm{PIdl}(O)/\mathrm{PIdl}(R) \\ \alpha F^\times O^\times &\mapsto \textit{class of } O\alpha O. \end{aligned} \tag{18.5.4}$$

If $\mathrm{PIdl}(O) \trianglelefteq \mathrm{Idl}(O)$ *is normal, then the isomorphism* (18.5.4) *induces a natural exact sequence*

$$\begin{aligned} 0 \to N_{B^\times}(O)/(F^\times O^\times) &\to \mathrm{Pic}_R(O) \to \mathrm{Idl}(O)/\mathrm{PIdl}(O) \to 0 \\ \alpha F^\times O^\times &\mapsto \textit{class of } O\alpha O. \end{aligned} \tag{18.5.5}$$

Proof. There is an isomorphism $N_{B^\times}(O)/O^\times \simeq \mathrm{PIdl}(O)$ by (18.5.2). The image of $F^\times \leq N_{B^\times}(O)$ in $\mathrm{PIdl}(O)$ under this map consists of two-sided ideals of the form OaO with $a \in F^\times$; we have $OaO = O$ if and only if $a \in O^\times$ if and only if $a \in R^\times$, so this image is isomorphic to the group $\mathrm{PIdl}(R)$ of principal fractional R-ideals via the map $aR \mapsto OaO$. The first isomorphism follows. The exact sequence (18.5.5) is then just rewriting the natural sequence

$$0 \to \mathrm{PIdl}(O)/\mathrm{PIdl}(R) \to \mathrm{Idl}(O)/\mathrm{PIdl}(R) \to \mathrm{Idl}(O)/\mathrm{PIdl}(O) \to 0.$$ □

Remark 18.5.6. The moral of Proposition 18.5.3 is that, unlike the commutative case where the two notions coincide, the two notions of "isomorphism classes of invertible bimodules" and "ideals modulo principal ideals" are in general different for a quaternion order. These notions coincide precisely when $N_{B^\times}(O)/F^\times \simeq O^\times/R^\times$, or equivalently (by the Skolem–Noether theorem) that every R-algebra automorphism of O is inner, which is to say $\operatorname{Aut}_R(O) = \operatorname{Inn}_R(O) = O^\times/R^\times$.

18.5.7. Unfortunately, the subgroup $\operatorname{PIdl}(O) \leq \operatorname{Idl}(O)$ need not be normal in general (Exercise 17.11), so statements like Proposition 18.5.3 depend on the order O having good structural properties. If O is a *maximal* order, then $\operatorname{Idl}(O)$ is abelian, so the result holds in this case.

In general, from the proof but using cosets one still obtains the equality

$$\#(\operatorname{Idl}(O)/\operatorname{PIdl}(O)) \cdot \#(N_{B^\times}(O)/(F^\times O^\times)) = \#\operatorname{Pic}_R(O). \tag{18.5.8}$$

Remark 18.5.9. If O, O' are connected, then $\operatorname{Pic}_R(O) \simeq \operatorname{Pic}_R(O')$ by 18.4.7 but this isomorphism need not respect the exact sequence (18.5.5). Each order O "balances" the contribution of this group between the normalizer $N_{B^\times}(O)/(F^\times O^\times)$ and the quotient $\operatorname{Idl}(O)/\operatorname{PIdl}(O)$—and these might be of different sizes for O'. We will return to examine more closely this structure in section 28.9, when strong approximation allows us to be more precise in measuring the discrepancy.

We conclude with an application to the structure of (right) class sets. We examine from Lemma 17.4.13 the fibers of the surjective map (17.4.14)

$$\begin{aligned} \operatorname{Cls} O &\to \operatorname{Typ} O \\ [I] &\mapsto \text{class of } O_{\mathsf{L}}(I). \end{aligned}$$

Refreshing our notation, let B be a central simple F-algebra and let $O \subset B$ an R-order.

Proposition 18.5.10. *The map $I \mapsto [I]$ induces a bijection*

$$\operatorname{PIdl}(O)\backslash \operatorname{Idl}(O) \leftrightarrow \{[I] \in \operatorname{Cls} O : O_{\mathsf{L}}(I) \simeq O\}.$$

Proof. Let O' be an order of the same type as O. Since (17.4.14) is surjective, there exists $[I] \in \operatorname{Cls} O$ such that $O_{\mathsf{L}}(I) \simeq O'$. We are free to replace O' by an isomorphic order, so we may suppose $O_{\mathsf{L}}(I) = O'$. For all $[I'] \in \operatorname{Cls} O$ with $O_{\mathsf{L}}(I') \simeq O'$ (running over the fiber), since $O_{\mathsf{L}}(\alpha I') = \alpha O' \alpha^{-1}$ for $\alpha \in B^\times$ we may suppose without loss of generality that the representative I' has $O_{\mathsf{L}}(I') = O'$.

We then define a map

$$\begin{aligned} \operatorname{PIdl}(O')\backslash \operatorname{Idl}(O') &\to \{[I'] \in \operatorname{Cls} O : O_{\mathsf{L}}(I') = O'\} \\ J' &\mapsto [J'I] \end{aligned} \tag{18.5.11}$$

The map is surjective, because if $J' = I'I^{-1}$ then $O_{\mathsf{L}}(J') = O_{\mathsf{R}}(J') = O$, so J' is a two-sided invertible O'-ideal. It is injective because if $[J'I] = [K'I]$ for $J', K' \in \operatorname{Idl}(O')$ then $K' = \alpha' J'$ with $\alpha' \in B^\times$, but further we need $O_{\mathsf{L}}(K') = \alpha' O' \alpha'^{-1} = O'$, so in fact $[J'I] = [K'I]$ if and only if $\alpha' \in N_{B^\times}(O')$, and the result then follows from Lemma 18.5.1. ∎

We have the following corollaries.

Corollary 18.5.12. *We have*

$$\#\operatorname{Cls} O = \sum_{[O']\in\operatorname{Typ} O} [\operatorname{Idl}(O') : \operatorname{PIdl}(O')] = \#\operatorname{Pic}_R O \sum_{[O']\in\operatorname{Typ} O} \frac{1}{z_{O'}}$$

where $z_{O'} = [N_{B^\times}(O') : F^\times O'^\times]$.

Proof. For the first equality, combine Lemma 17.4.13 and Proposition 18.5.10, computing the size of the fibers. For the second, substitute (18.5.8) and use 18.4.7. □

Corollary 18.5.13. *Let* O_i *be representatives of* $\operatorname{Typ} O$. *For each* i, *let* I_i *be a connecting* O_i, O*-ideal, and let* $J_{i,j}$ *be representatives of* $\operatorname{PIdl}(O_i)\backslash \operatorname{Idl}(O_i)$. *Then the set* $\{J_{i,j}I_i\}_{i,j}$ *is a complete set of representatives for* $\operatorname{Cls} O$.

Proof. We choose representatives and take the fibers of the map (17.4.14). □

Remark 18.5.14. When $\operatorname{PIdl}(O) \trianglelefteq \operatorname{Idl}(O)$, then in Proposition 18.5.10 we have written the class set $\operatorname{Cls} O$ as a disjoint union of abelian groups. The fact that the bijection is noncanonical is due to the fact that we choose a connecting ideal, so without making choices we obtain only a disjoint union of principal homogeneous spaces (i.e., torsors) under the groups $\operatorname{PIdl}(O')\backslash \operatorname{Idl}(O')$.

Exercises

Unless otherwise specified, let R be a Dedekind domain with field of fractions $F = \operatorname{Frac} R$, let B be a simple finite-dimensional F-algebra, and let $O \subseteq B$ be an R-order.

1. Show that the following are equivalent:

   (i) O is a maximal R-order;
   (ii) $O_{\mathsf{L}}(I) = O_{\mathsf{R}}(I) = O$ for all fractional two-sided O-ideals I; and
   (iii) $O_{\mathsf{L}}(I) = O_{\mathsf{R}}(I) = O$ for all two-sided O-ideals $I \subseteq O$.

2. Show that the zero ideal is a prime ideal of O.
3. Let $J \subseteq O$ be a nonzero two-sided ideal of O in the ring-theoretic sense: J is an additive subgroup closed under left and right multiplication by O. Show that J is an R-lattice.
4. Let R be a DVR with maximal ideal $\mathfrak{p}$, and let $O = \begin{pmatrix} R & R \\ \mathfrak{p} & R \end{pmatrix} \subseteq B = \mathrm{M}_2(F)$. Show that the two-sided ideal $\mathfrak{p}\,\mathrm{M}_2(R) \subseteq O$ is *not* a prime ideal.
5. Let $R := \mathbb{Z}[\sqrt{-6}]$ and $F := \mathbb{Q}(\sqrt{-6})$. Let $B := (2, \sqrt{-6} \mid F)$.

   (a) Show that $2R = \mathfrak{p}_2^2$ and $3R = \mathfrak{p}_3^2$ for primes $\mathfrak{p}_2, \mathfrak{p}_3 \subseteq R$.
   (b) Show that $\operatorname{Ram}(B) = \{\mathfrak{p}_2, \mathfrak{p}_3\}$.
   (c) Let O be a maximal order in B. Show that there is a unique two-sided ideal P_2 such that $P_2^2 = \mathfrak{p}_2 O$.

(d) Prove that $[P_2] \in \mathrm{Pic}_R(O)$ has order 4, and conclude that the sequence (18.4.9) does not split.
(e) Show that we may take

$$O = R + \mathfrak{p}_2^{-1}(\sqrt{-6} + i) + Rj + \mathfrak{p}_2^{-1}(\sqrt{-6} + i)j$$

as the maximal order, and then that I is generated by i and $\sqrt{-6}ij/2$, and finally that $I^2 = (\sqrt{-6} + i)/2$.

6. Let $B = \mathrm{M}_n(F)$ with $n \geq 2$, let $O = \mathrm{M}_n(R)$, let $\mathfrak{p} \subseteq R$ be prime with $k = R/\mathfrak{p}$, and let $O(\mathfrak{p}) = R + \mathfrak{p}O$.

(a) Show that $O(\mathfrak{p})$ is an order of reduced discriminant $\mathfrak{p}^3$.
(b) Show that $O^\times \simeq \mathrm{GL}_n(R)$ normalizes $O(\mathfrak{p}) \subseteq O$, so that

$$O(\mathfrak{p})^\times \trianglelefteq O^\times \simeq \mathrm{GL}_n(R),$$

and that the map

$$\begin{aligned} O^\times &\hookrightarrow \mathrm{Idl}(O) \\ \gamma &\mapsto O\gamma = \gamma O \end{aligned}$$

induces an injective group homomorphism $\mathrm{PGL}_n(k) \hookrightarrow \mathrm{Idl}(O)$. Conclude that $\mathrm{Idl}(O)$ is not an abelian group.

7. Show that Theorem 18.3.4 holds more generally for B a *semi*simple F-algebra (but still $O \subseteq B$ maximal). *[Hint: Decompose B into a product of simple F-algebras.]*
8. Let O be maximal, and let $P_1, \dots, P_r \subseteq O$ be distinct prime two-sided ideals. Let

$$I := \prod_{i=1}^r P_i^{e_i} \quad \text{and} \quad J := \prod_{i=1}^r P_i^{f_i}$$

with $e_i, f_i \in \mathbb{Z}$.

(a) Prove that $I \subseteq O$ if and only if $e_i \geq 0$ for all $i = 1, \dots, n$, and in this case there is a ring isomorphism

$$O/I \simeq \bigoplus_{i=1}^r O/P_i^{e_i}.$$

(b) Prove that $I \supseteq J$ if and only if $e_i \leq f_i$ for all i.
(c) Show $I + J = \prod_{i=1}^r P_i^{\min(e_i, f_i)}$ and $I \cap J = \prod_{i=1}^r P_i^{\max(e_i, f_i)}$.

▹ 9. Prove Lemma 18.4.4: Show that fractional two-sided O-ideals $I, J \subseteq B$ are isomorphic as O-bimodules over R if and only if there exists $a \in F^\times$ such that $J = aI$. *[Hint: Peek at Lemma 19.5.1.]*

10. Let $K \supseteq F$ be a finite, separable extension and let S be the integral closure of R in K. Show that the map $I \mapsto I \otimes_R S$ defines a group homomorphism $\mathrm{Pic}\, O \to \mathrm{Pic}(O \otimes_R S)$.

This chapter is licensed under the terms of the Creative Commons Attribution-NonCommercial 4.0 International License (http://creativecommons.org/licenses/by-nc/4.0/), which permits any noncommercial use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license and indicate if changes were made.

The images or other third party material in this chapter are included in the chapter's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the chapter's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.

Chapter 19

Brandt groupoids

In this chapter, we study the relationship between multiplication and classes of quaternion ideals.

19.1 ▹ Composition laws and ideal multiplication

To guide our investigations, we again appeal to the quadratic case. Let $d \in \mathbb{Z}$ be a nonsquare discriminant. A subject of classical interest was the set of **integral primitive binary quadratic forms** of discriminant d, namely

$$Q(d) = \{ax^2 + bxy + cy^2 : a, b, c \in \mathbb{Z}, b^2 - 4ac = d, \text{ and } \gcd(a, b, c) = 1\}.$$

Of particular interest to early number theorists (Fermat, Legendre, Lagrange, and Gauss) was the set of primes represented by a quadratic form $Q \in Q(d)$; inquiries of this nature proved to be quite deep, giving rise to the law of quadratic reciprocity and the beginnings of the theory of complex multiplication and class field theory.

An invertible, oriented change of variables on a quadratic form $Q \in Q(d)$ does not alter the set of primes represented, so one is naturally led to study the equivalence classes of quadratic forms under the (right) action of the group $\mathrm{SL}_2(\mathbb{Z})$ given by

$$(Q \mid g)(x, y) = Q((x, y) \cdot g) \quad \text{for} \quad g \in \mathrm{SL}_2(\mathbb{Z}). \tag{19.1.1}$$

The set $\mathrm{Cl}(d)$ of $\mathrm{SL}_2(\mathbb{Z})$-classes of forms in $Q(d)$ is finite, by reduction theory (see section 35.2): every form in $Q(d)$ is equivalent under the action of $\mathrm{SL}_2(\mathbb{Z})$ to a unique **reduced** form, of which there are only finitely many. To study this finite set, Gauss defined a **composition law** on $\mathrm{Cl}(d)$, giving $\mathrm{Cl}(d)$ the structure of an abelian group by an explicit formula. Gauss's composition law on binary quadratic forms can be understood using $2 \times 2 \times 2$ Rubik's cubes, by a sublime result of Bhargava [Bha2004a].

Today, we see this composition law as a consequence of a natural bijection between $\mathrm{Cl}(d)$ and a set equipped with an obvious group structure. Let $S = S(d)$ be the quadratic ring of discriminant d. Define the **narrow class group** $\mathrm{Cl}^+(S)$ as

the group of invertible fractional ideals of S under multiplication

© The Author(s) 2021

J. Voight, *Quaternion Algebras*, Graduate Texts in Mathematics 288,
https://doi.org/10.1007/978-3-030-56694-4_19

modulo

the subgroup of nonzero principal fractional ideals
generated by a totally positive element

(i.e., one that is positive in every embedding into $\mathbb{R}$, so if $d < 0$ then this is no condition). (Alternatively, $\mathrm{Cl}^+(S)$ can be thought of as the group of isomorphism classes of *oriented*, invertible S-modules, under a suitable notion of orientation.) Then there is a bijection between $\mathrm{Cl}(d)$ and $\mathrm{Cl}^+(S)$: explicitly, to the class of the quadratic form $Q = ax^2 + bxy + cy^2 \in \mathcal{Q}(d)$, we associate the class of the ideal

$$\mathfrak{a} = a\mathbb{Z} + \left(\frac{-b+\sqrt{d}}{2}\right)\mathbb{Z} \subset S(d). \tag{19.1.2}$$

Conversely, the quadratic form is recovered as the norm form on $K = \mathbb{Q}(\sqrt{d})$ restricted to $\mathfrak{a}$:

$$\mathrm{Nm}_{K/\mathbb{Q}}\left(ax + \frac{-b+\sqrt{d}}{2}y\right) = ax^2 + bxy + cy^2 \tag{19.1.3}$$

where $c = \frac{b^2-d}{4a} \in \mathbb{Z}$.

Much of the same structure can be found in the quaternionic case, with several interesting twists. It was Brandt who first asked if there was a composition law for (integral, primitive) quaternary quadratic forms: it would arise naturally from some kind of multiplication of ideals in a quaternion order, with the analogous bijection furnished by the reduced norm form. Brandt started writing on composition laws for quaternary quadratic forms in 1913 [Bra13], tracing the notion of composition back to Hermite, who observed a kind of multiplication law (bilinear substitution) for quaternary forms $x_0^2 + F(x_1, x_2, x_3)$ in formulas of Euler and Lagrange. He continued on this note during the 1920s [Bra24, Bra25, Bra28, Bra37], when it became clear that quaternion algebras was the right framework to place his composition laws; in 1943, he developed this theme significantly [Bra43] and defined his *Brandt matrices* (that will figure prominently in Chapter 41.

However, in the set of invertible lattices in B under compatible product, one cannot always multiply! However, this set has the structure of a **groupoid**: a nonempty set with an inverse function and a partial product that satisfies the associativity, inverse, and identity properties whenever they are defined. Groupoids now figure prominently in category theory (a groupoid is equivalently a small category in which every morphism is an isomorphism) and many other contexts; see Remark 19.3.12.

Organizing lattices by their left and right orders, which by definition are connected and hence in the same genus, we define

$$\mathrm{Brt}(\mathcal{O}) = \{I : I \subset B \text{ invertible } R\text{-lattice and } \mathcal{O}_{\mathsf{L}}(I), \mathcal{O}_{\mathsf{R}}(I) \in \mathrm{Gen}\,\mathcal{O}\}; \tag{19.1.4}$$

visibly, $\mathrm{Brt}(\mathcal{O})$ depends only on the genus of $\mathcal{O}$. Organizing lattices according to the genus of orders is sensible: after all, we only apply the composition law to binary quadratic forms of the same discriminant, and in the compatible product we see precisely those classes whose left and right orders are connected. In other words, the set of invertible lattices in the quadratic field $K = \mathbb{Q}(\sqrt{d})$ has the structure of a groupoid if we multiply only those lattices with the same multiplicator ring.

Theorem 19.1.5. *Let B be a quaternion algebra over $\mathbb{Q}$ and let $O \subset B$ be an order. Then the set* $\mathrm{Brt}(O)$ *has the structure of a groupoid under compatible product.*

We call $\mathrm{Brt}(O)$ the **Brandt groupoid** of (the genus of) O.

We now consider *classes* of lattices. A lattice $I \subset B$ has the structure of a $O_{\mathsf{L}}(I), O_{\mathsf{R}}(I)$-bimodule. Two invertible lattices I, J with the same left and right orders $O_{\mathsf{L}}(I) = O_{\mathsf{L}}(J)$ and $O_{\mathsf{R}}(I) = O_{\mathsf{R}}(J)$ are isomorphic as bimodules if and only if there exists $a \in \mathbb{Q}^\times$ such that $J = aI$. Accordingly, we say two lattices $I, J \subset B$ are **homothetic** if there exists $a \in \mathbb{Q}^\times$ such that $J = aI$.

For connected orders $O, O' \subset B$, we define

$$\mathrm{Pic}(O, O') := \{[I] : I \subset B \text{ invertible and } O_{\mathsf{L}}(I) = O \text{ and } O_{\mathsf{R}}(I) = O'\} \tag{19.1.6}$$

to be the set of homothety classes of lattices with left order O and right order O', or equivalently the set of isomorphism classes of O, O'-bimodules over R. Restricting to the subset of lattices with $O = O'$, and the lattices $I \subset B$ are O-bimodules, we recover $\mathrm{Pic}(O, O) = \mathrm{Pic}\, O$ the Picard group from the previous chapter.

Now let $O \subset B$ be an order and let O_i be representative orders for the type set $\mathrm{Typ}\, O$. Let

$$\mathrm{BrtCl}\, O := \bigsqcup_{i,j} \mathrm{Pic}(O_i, O_j). \tag{19.1.7}$$

Theorem 19.1.8. *Let B be a quaternion algebra over $\mathbb{Q}$ and let $O \subset B$ be an order. Then the set* $\mathrm{BrtCl}\, O$ *has the structure of a groupoid that, up to isomorphism, is independent of the choice of the orders O_i.*

In particular, $\mathrm{BrtCl}\, O$ depends only on the genus of O. We call the set $\mathrm{BrtCl}\, O$ the **Brandt class groupoid** of (the genus of) O.

Returning to quadratic forms, to each R-lattice I with $\mathrm{nrd}(I) = a\mathbb{Z}$ and $a > 0$, we associate the quadratic form

$$\begin{aligned} \mathrm{nrd}_I &: I \to \mathbb{Z} \\ \mathrm{nrd}_I(\mu) &= \mathrm{nrd}(\mu)/a \end{aligned}$$

Alternatively, up to similarity we can just take the quadratic module $\mathrm{nrd}|_I : I \to \mathrm{nrd}(I)$ remembering that the quadratic form takes values in $\mathrm{nrd}(I)$. The discriminant of an invertible lattice $I \subset B$ is equal to the common discriminant N^2 of the genus of its left or right order. The quadratic forms nrd_I are all locally similar, respecting the canonical orientation 5.6.7 on B. Therefore, there is a map

$$\begin{aligned} \mathrm{BrtCl}\, O &\to \left\{ \begin{array}{c} \text{Quaternary quadratic forms over } \mathbb{Z} \\ \text{locally similar to } \mathrm{nrd}|_O \\ \text{up to oriented similarity} \end{array} \right\} \\ [I] &\mapsto \mathrm{nrd}_I \end{aligned}$$

is (well-defined and) surjective. Unfortunately, this map is not injective (a reflection of the lack of a natural quotient groupoid homomorphism): the Brandt class is a kind of rigidification of the oriented similarity class. Nevertheless, Theorem 19.1.8 can be viewed as a generalization of Gauss composition of binary quadratic forms, defining a partial composition law on (rigidified) classes of quaternary quadratic forms.

19.2 Example

Consider the quaternion algebra $B := \left(\dfrac{-2,-37}{\mathbb{Q}}\right)$ with standard basis $1, i, j, k = ij$, and the maximal order $\mathcal{O}$ of reduced discriminant 37 defined by

$$\mathcal{O} := \mathbb{Z} + i\mathbb{Z} + \frac{1+i+j}{2}\mathbb{Z} + \frac{2+i+k}{4}\mathbb{Z}. \tag{19.2.1}$$

The type set $\operatorname{Typ}\mathcal{O}$ of orders connected to $\mathcal{O}$ has exactly two isomorphism classes, represented by $\mathcal{O}_1 = \mathcal{O}$ and

$$\mathcal{O}_2 := \mathbb{Z} + 3i\mathbb{Z} + \frac{3-7i+j}{6}\mathbb{Z} + \frac{2-3i+k}{4}\mathbb{Z}.$$

These orders are connected by the $\mathcal{O}_2, \mathcal{O}_1$-connecting ideal

$$I := 3\mathbb{Z} + 3i\mathbb{Z} + \frac{3-i+j}{2}\mathbb{Z} + \frac{2+3i-k}{4}\mathbb{Z} = 3\mathcal{O} + \frac{3-i+j}{2}\mathcal{O}.$$

There are isomorphisms

$$\operatorname{Pic}(\mathcal{O}) \simeq \operatorname{Pic}(\mathcal{O}_2) \simeq \mathbb{Z}/2\mathbb{Z}$$

with the nontrivial class in $\operatorname{Pic}(\mathcal{O}_1)$ represented by the principal two-sided ideal $J_1 = j\mathcal{O} = \mathcal{O}j$ with $j \in N_{B^\times}(\mathcal{O})$, and the nontrivial class in $\operatorname{Pic}(\mathcal{O}_2)$ represented by the nonprincipal (but invertible) ideal

$$J_2 := IJ_1I^{-1} = 37\mathcal{O}_2 + \frac{111-259i+j}{6}\mathcal{O}_2.$$

In particular,

$$\operatorname{Cls}_{\mathsf{R}}(\mathcal{O}_1) = \{[\mathcal{O}_1], [I], [J_2I]\} \quad \text{and} \quad \operatorname{Cls}_{\mathsf{R}}(\mathcal{O}_2) = \{[\mathcal{O}_2], [J_2], [\overline{I}]\},$$

with $[J_1] = [\mathcal{O}_1]$.

We can visualize this groupoid as a graph as in Figure 19.2.2, with directed edges for multiplication:

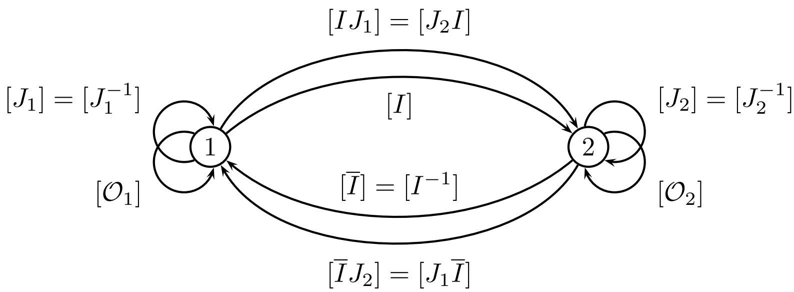


Figure 19.2.2: $\operatorname{BrtCl}\mathcal{O}$, for $\operatorname{discrd}\mathcal{O} = 37$

The Brandt class groupoid

$$\mathrm{BrtCl}\, O = \mathrm{Pic}(O_1) \sqcup \mathrm{Cl}(O_1, O_2) \sqcup \mathrm{Cl}(O_2, O_1) \sqcup \mathrm{Pic}(O_2)$$

has $2 + 4 + 4 + 2 = 12$ elements; it is generated as a groupoid by the elements $[J_1], [J_2], [I]$, with relations

$$[J_1]^2 = [O_1], \quad [J_2]^2 = [O_2], \quad [J_2][I] = [I][J_1].$$

Restricting the reduced norm to these lattices, we obtain classes of quaternary quadratic forms of discriminant 37^2:

$$\begin{aligned}
\mathrm{nrd}_{O_1} &= t^2 + ty + tz + 2x^2 + xy + 2xz + 5y^2 + yz + 10z^2 \\
\mathrm{nrd}_{O_2} &= t^2 + tx + tz + 4x^2 - xy + 4xz + 5y^2 + 2yz + 6z^2 \\
\mathrm{nrd}_{I} &= 3t^2 - tx + ty + tz + 3x^2 - 3xy - xz + 4y^2 - yz + 5z^2 \\
\mathrm{nrd}_{\overline{I}} &= 3t^2 + tx - ty - tz + 3x^2 - 3xy - xz + 4y^2 - yz + 5z^2 \\
\mathrm{nrd}_{J_2} &= 2t^2 - tx + ty + 2x^2 - 2xy + xz + 3y^2 + 2yz + 10z^2
\end{aligned}$$

The quadratic forms nrd_I and $\mathrm{nrd}_{\overline{I}}$ are isometric but not by an oriented isometry.

19.3 Groupoid structure

We begin with some generalities on groupoids.

Definition 19.3.1. A **partial function** $f\colon X \to Y$ is a function defined on a subset of the domain X.

Definition 19.3.2. A **groupoid** G is a set with a unary operation $^{-1} : G \to G$ and a partial function $* : G \times G \to G$ such that $*$ and $^{-1}$ satisfy the associativity, inverse, and identity properties (as in a group) whenever they are defined:

(a) [Associativity] For all $a, b, c \in G$ such that $a * b$ is defined and $(a * b) * c$ is defined, both $b * c$ and $a * (b * c)$ are defined and

$$(a * b) * c = a * (b * c).$$

(b) [Inverses] For all $a \in G$, there exists $a^{-1} \in G$ such that $a * a^{-1}$ and $a^{-1} * a$ are defined (but not necessarily equal).

(c) [Identity] For all $a, b \in G$ such that $a * b$ is defined, we have

$$(a * b) * b^{-1} = a \quad \text{and} \quad a^{-1} * (a * b) = b. \tag{19.3.3}$$

A **homomorphism** $\phi\colon G \to G'$ of groupoids is a map satisfying

$$\phi(a * b) = \phi(a) * \phi(b)$$

for all $a, b \in G$.

19.3.4. Let G be a groupoid. Then the products in the identity law (19.3.3) are defined by the associative and inverse laws, and it follows that $e = a * a^{-1}$, the **left identity** of a, and $f = a^{-1} * a$ the corresponding **right identity** of a, satisfy $e * a = a = a * f$ for all $a \in G$. (We may have that $e \neq f$, i.e., the left and right identities for $a \in G$ disagree.) The right identity of $a \in G$ is the left identity of $a^{-1} \in G$, so we call the set

$$\{e = a * a^{-1} : a \in G\}$$

the set of **identity elements** in G.

19.3.5. Equivalently, a groupoid is a small category (the class of objects in the category is a set) such that every morphism is an isomorphism: given a groupoid, we associate the category whose objects are the elements of the set $S := \{e = a * a^{-1} : a \in G\}$ of identity elements in G and the morphisms between $e, f \in S$ are the elements $a \in G$ such that $e * a$ and $a * f$ are defined (see Proposition 19.3.9 below). Conversely, to a category in which every morphism is an isomorphism, we associate the groupoid whose underlying set is the union of all morphisms under inverse and composition.

Example 19.3.6. The set of homotopy classes of paths in a topological space X forms a groupoid under composition: the paths $\gamma_1, \gamma_2 : [0, 1] \to X$ can be composed to a path $\gamma_2 \circ \gamma_1 : [0, 1] \to X$ if and only if $\gamma_2(0) = \gamma_1(1)$.

Example 19.3.7. A disjoint union of groups is a groupoid, with the product defined if and only if the elements belong to the same group; the set of identities is canonically in bijection with the index set of the disjoint union.

19.3.8. Let G be a groupoid and let $e, f \in G$ be identity elements. We say that e is **connected** to f if there exists $a \in G$ such that a has left identity e and right identity f. The relation of being connected defines an equivalence relation on the set of identity elements in G, and the resulting equivalence classes are called **connected components** of G. We say G is **connected** if all identity elements $e, f \in G$ are connected; connected components of a groupoid are connected.

Viewing the groupoid G as a small category as in 19.3.5, we say two objects are connected if there exists a morphism between them, and the category is connected if every two objects are connected.

If $e \in G$ is an identity element in a groupoid G, then the set of elements $a \in G$ with left and right identity equal to e has the structure of a group; for the associated category, this is the automorphism group of the object. More generally, the following structural result holds.

Proposition 19.3.9. *Let G be a connected groupoid, and let e, f be identity elements in G. Let*

$$G(e, f) := \{a \in G : e * a \text{ and } a * f \text{ are defined}\}. \tag{19.3.10}$$

Then the following statements hold.

(a) *The set $G(e, e)$ is a group under $*$.*
(b) *There is a (noncanonical) isomorphism $G(e, e) \simeq G(f, f)$.*

(c) *The set* $G(e, f)$ *is a principal homogeneous space for* $G(e, e) \simeq G(f, f)$.

Proof. The set $G(e, e)$ is nonempty, has the identity element $e \in G$, and if $a \in G(e, e)$ then $a * a^{-1} = a^{-1} * a = e$. If e, f are identity elements, since G is connected there exists $a \in G(e, f)$, so $a^{-1} \in G(f, e)$ and the map $G(e, e) \to G(f, f)$ by $x \mapsto a * x * a^{-1}$ is an isomophism of groups. Similarly, the set $G(e, f)$ has a right, simply transitive action of $G(e, e)$ under right multiplication by $*$. ☐

19.3.11. The moral of Proposition 19.3.9 is that the only two interesting invariants of a connected groupoid are the number of identity elements (objects in the category) and the group of elements with a common left and right identity (the automorphism group of every one of the objects). A connected groupoid is determined up to isomorphism of groupoids by these two properties.

Remark 19.3.12. After seeing its relevance in the context of composition of quaternary forms, Brandt set out general axioms for his notion of a groupoid [Bra27, Bra40]. (Brandt's original definition of groupoid is now called a *connected* groupoid.) This notion has blossomed into an important structure in mathematics that sees quite general use, especially in homotopy theory and category theory. It is believed that the groupoid axioms influenced the work of Eilenberg–Mac Lane [EM45] in the first definition of a category: see e.g., Brown [Bro87] for a survey, Bruck [Bruc71] for context in the theory of binary structures, as well as the article by Weinstein [Wein96].

Groupoids exhibit many facets of mathematics, arising naturally in functional analysis (C^*-algebras) and group representations, as Figure 19.3.13 indicates (appearing in Williams [Will2001, p. 21], and attributed to Arlan Ramsay).

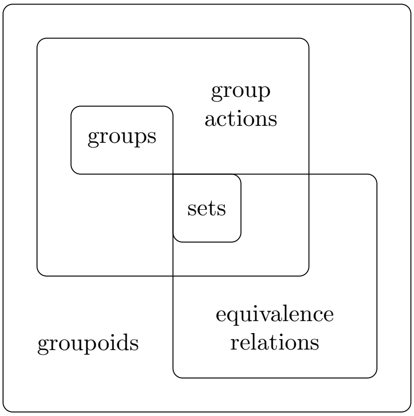


Figure 19.3.13: Groupoids, as they relate to other mathematical objects

(In this diagram, for example, a set X is a groupoid with only the multiplications $x * x = x$ for $x \in X$. The corner between sets and groups can be explained by a set with one element which can be made into a group in a unique way.)

19.4 Brandt groupoid

Let R be a Dedekind domain with field of fractions F and let B be a quaternion algebra over F.

Proposition 19.4.1. *The set of invertible R-lattices in B is a groupoid under inverse and compatible product; the R-orders in B are the identity elements in this groupoid.*

Proof. For the associative law, suppose I, J, K are invertible R-lattices with IJ and $(IJ)K$ compatible products. Then $\mathcal{O}_{\mathsf{R}}(I) = \mathcal{O}_{\mathsf{L}}(J) = \mathcal{O}_{\mathsf{L}}(JK)$ and $\mathcal{O}_{\mathsf{R}}(IJ) = \mathcal{O}_{\mathsf{R}}(J) = \mathcal{O}_{\mathsf{L}}(K)$ by Lemma 16.5.11, so the products JK and $I(JK)$ are compatible. Multiplication is associative in B, and it follows that $I(JK) = (IJ)K$. Inverses exist exactly because we restrict to the invertible lattices.

The law of identity holds as follows: if I, J are invertible R-lattices such that IJ is a compatible product, then $(IJ)J^{-1}$ is a compatible product since $\mathcal{O}_{\mathsf{R}}(IJ) = \mathcal{O}_{\mathsf{R}}(J) = \mathcal{O}_{\mathsf{L}}(J^{-1})$, and by associativity

$$(IJ)J^{-1} = I(JJ^{-1}) = I\mathcal{O}_{\mathsf{L}}(J) = I\mathcal{O}_{\mathsf{R}}(I) = I,$$

with a similar argument on the left. If I is an invertible R-lattice, then $II^{-1} = \mathcal{O}_{\mathsf{L}}(I)$ is an R-order in B, and every R-order $\mathcal{O}$ arises by taking $I = \mathcal{O}$ itself, so the R-orders are the identity elements in the groupoid. ◻

Lemma 19.4.2. *The connected components of the groupoid of invertible R-lattices in B are identified by the genus of the (left or) right order, and the group defined on such a component corresponding to an order $\mathcal{O}$ is* $\operatorname{Idl}(\mathcal{O})$*, the group of invertible two-sided $\mathcal{O}$-ideals.*

Proof. By Proposition 19.4.1, the identity elements correspond to orders, and two orders are connected if and only if there is a (invertible, equivalently locally principal) connecting ideal if and only if they are in the same genus, as in section 17.4. The second statement follows immediately. ◻

As a consequence of Lemma 19.4.2, the subset of R-lattices whose (left or) right order belong to a specified genus of orders is a connected subgroupoid.

Definition 19.4.3. Let $\mathcal{O} \subseteq B$ be an R-order. The **Brandt groupoid** of (the genus of) $\mathcal{O}$ is

$$\operatorname{Brt}(\mathcal{O}) = \{I : I \subset B \text{ invertible } R\text{-lattice and } \mathcal{O}_{\mathsf{L}}(I), \mathcal{O}_{\mathsf{R}}(I) \in \operatorname{Gen}\mathcal{O}\}.$$

In the next section, we consider a variant that considers classes of lattices, giving rise to a finite groupoid.

19.5 Brandt class groupoid

We now organize lattices up to isomorphism as bimodules for their left and right orders.

Lemma 19.5.1. *Let $I, J \subset B$ be lattices with $\mathcal{O}_{\mathsf{L}}(I) = \mathcal{O}_{\mathsf{L}}(J) = \mathcal{O}$ and $\mathcal{O}_{\mathsf{R}}(I) = \mathcal{O}_{\mathsf{R}}(J) = \mathcal{O}'$. Then I is isomorphic to J as $\mathcal{O}, \mathcal{O}'$-bimodules if and only if there exists $a \in F^\times$ such that $J = aI$.*

Proof. We have $F = Z(B)$. If $J = aI$ with $a \in F^\times$, then multiplication by a gives an R-module isomorphism $I \to J$ that commutes with the left and right actions and so defines a $\mathcal{O}, \mathcal{O}'$-bimodule isomorphism.

Conversely, suppose that $\phi\colon I \xrightarrow{\sim} J$ is a $\mathcal{O}, \mathcal{O}'$-bimodule isomorphism. Then $\phi(\mu\alpha\nu) = \mu\phi(\alpha)\nu$ for all $\alpha \in I$ and $\mu, \nu \in \mathcal{O}$. Extending scalars to B, we obtain a B-bimodule isomorphism $\phi\colon IF = B \to JF = B$. Let $\phi(1) = \beta$. Then for all $\alpha \in B$, we have $\phi(\alpha) = \phi(1)\alpha = \beta\alpha$; but by the same token, $\phi(\alpha) = \alpha\beta$ for all $\alpha \in B$, so $\beta \in Z(B) = F$. ◻

Definition 19.5.2. Let $I, J \subseteq B$ be R-lattices. We say that I is **homothetic** to J if there exists $a \in F^\times$ such that $J = aI$.

Homothety defines an equivalence relation, and we let $[I]$ denote the homothety class of an R-lattice I. The left and right order of a homothety class is well-defined.

19.5.3. The set of homothety classes of invertible R-lattices $I \subseteq B$ has the structure of a groupoid under compatible product, since the compatible product $[IJ]$ is well-defined: if $I' = aI$ and $J' = bJ$ with $a, b \in F^\times$, then $[I'J'] = [abIJ] = [IJ]$ since a, b are central.

The map which takes an invertible lattice to its homothety class yields a surjective homomorphism of groupoids. Taking connected components we obtain a connected groupoid associated to a (genus of an) R-order $\mathcal{O}$. Recalling 19.3.11, we note that the group at an R-order $\mathcal{O}$ is $\operatorname{Pic}_R(\mathcal{O})$, but there are still infinitely many orders (objects in the category).

In order to whittle down to a finite groupoid, we fix representatives of the type set, and make the following definitions.

19.5.4. For R-orders $\mathcal{O}, \mathcal{O}' \subseteq B$, let

$$\operatorname{Pic}_R(\mathcal{O}, \mathcal{O}') = \{[I] : I \subset B \text{ invertible and } \mathcal{O}_{\mathsf{L}}(I) = \mathcal{O}, \mathcal{O}_{\mathsf{R}}(I) = \mathcal{O}'\}$$

be the set of homothety classes of R-lattices in B with left order $\mathcal{O}$ and right order $\mathcal{O}'$; equivalently, by Lemma 19.5.1, $\operatorname{Pic}(\mathcal{O}, \mathcal{O}')$ is the set of isomorphism classes of invertible $\mathcal{O}, \mathcal{O}'$-bimodules over R. In particular, $\operatorname{Pic}_R(\mathcal{O}) = \operatorname{Pic}_R(\mathcal{O}, \mathcal{O})$.

We have $\operatorname{Pic}_R(\mathcal{O}, \mathcal{O}') \neq \emptyset$ if and only if $\mathcal{O}$ is connected to $\mathcal{O}'$.

Let $\mathcal{O} \subset B$ be an order and let $\mathcal{O}_i$ be representative orders for the type set $\operatorname{Typ} \mathcal{O}$. We define

$$\operatorname{BrtCl} \mathcal{O} := \bigsqcup_{i,j} \operatorname{Pic}_R(\mathcal{O}_i, \mathcal{O}_j).$$

Theorem 19.5.5. *Let R be a Dedekind domain with field of fractions F, and let B be a quaternion algebra over F. Let $O \subset B$ be an order. Then the set* $\operatorname{BrtCl} O$ *has the structure of a* finite *groupoid that, up to isomorphism, is independent of the choice of the orders O_i.*

In particular, by Theorem 19.5.5 $\operatorname{BrtCl} O$ depends only on the genus of O up to groupoid isomorphism. We call the set $\operatorname{BrtCl} O$ the **Brandt class groupoid** of (the genus of) O.

Proof. The groupoid structure is compatible multiplication, with

$$\operatorname{Pic}_R(O_i, O_j)\operatorname{Pic}_R(O_j, O_k) \subseteq \operatorname{Pic}_R(O_i, O_k)$$

for all i, j, k; in other words, $\operatorname{BrtCl} O$ is a connected subgroupoid of the groupoid of homothety classes of R-lattices 19.5.4.

The groupoid is finite, by 19.3.11: the type set $\operatorname{Typ} O$ is finite by Main Theorem 17.7.1 and $\operatorname{Pic}_R(O)$ is finite by Proposition 18.4.10. Explicitly, if $[I_{ij}] \in \operatorname{Pic}_R(O_i, O_j)$ then the map

$$\begin{aligned} \operatorname{Pic}_R(O) \simeq \operatorname{Pic}_R(O_i) &\to \operatorname{Pic}_R(O_i, O_j) \\ [I] &\mapsto [II_{ij}] \end{aligned}$$

is a bijection of sets, just as in the proof of Proposition 19.3.9. Therefore

$$\#\operatorname{BrtCl} O = \#\operatorname{Pic}_R(O)\#\operatorname{Typ} O. \tag{19.5.6}$$

Finally, this subgroupoid is independent of the choices of the orders O_i as follows: all other choices correspond to $O_i' = \alpha_i O_i \alpha_i^{-1}$ with $\alpha_i \in B^\times$, and the induced maps

$$\begin{aligned} \operatorname{Pic}_R(O_i, O_j) &\to \operatorname{Pic}(O_i', O_j') \\ [I] &\mapsto [\alpha_i I \alpha_j^{-1}] = [I'] \end{aligned}$$

together give an isomorphism of groupoids, since

$$[I'J'] = [\alpha_i I \alpha_j^{-1} \alpha_j J \alpha_k^{-1}] = [\alpha_i I J \alpha_k^{-1}]$$

for all $[I] \in \operatorname{Pic}_R(O_i, O_j)$ and $[J] \in \operatorname{Pic}_R(O_j, O_k)$. □

Remark 19.5.7. Unfortunately, there is *not* in general a natural equivalence relation on $\operatorname{Brt}(O)$ giving rise to a quotient groupoid homomorphism $\operatorname{Brt}(O) \to \operatorname{BrtCl} O$. Rather, we find that $\operatorname{BrtCl} O$ is naturally a subgroupoid of $\operatorname{Brt}(O)$.

Turning to the invariants 19.3.11, we see that the Brant class groupoid $\operatorname{BrtCl} O$ encodes two things: the group $\operatorname{Pic}_R(O)$ and the type set $\operatorname{Typ} O$.

Remark 19.5.8. The modern theory of Brandt composition was investigated by Kaplansky [Kap69] and generalized to Azumaya quaternion algebras over commutative rings by Kneser–Knus–Ojanguren–Parimala–Sridharan [KKOPS86] for a generalization of the composition law to Azumaya algebras over rings.

19.6 Quadratic forms

We now connect the Brandt class groupoid to quadratic forms. For simplicity, we suppose $\operatorname{char} F \neq 2$ throughout this section.

19.6.1. We begin by recalling Proposition 4.5.17: for the quaternary quadratic form $\operatorname{nrd}\colon B \to F$, every oriented similarity of nrd is of the form

$$\begin{aligned} B &\mapsto B \\ x &\mapsto \alpha x \beta^{-1} \end{aligned}$$

with $\alpha, \beta \in B$ (in particular, respecting the canonical orientation 5.6.7 of B); the similitude factor of such a map is $u = \operatorname{nrd}(\alpha)/\operatorname{nrd}(\beta)$.

Let $I \subset B$ be a projective R-lattice.

19.6.2. Generalizing Exercise 10.2, the reduced norm restricts to give a quadratic form on I. We are given that I is projective of rank 4 as an R-module. Therefore the map

$$\operatorname{nrd}_I : I \to L = \operatorname{nrd}(I)$$

is a quaternary quadratic module over R.

If $J \subset B$ is another projective R-lattice, and f is an oriented similarity from nrd_I to nrd_J, then extending scalars by F we obtain a oriented self-similarity of $\operatorname{nrd}\colon B \to B$; by 19.6.1, we conclude that $J = \alpha I \beta^{-1}$ for some $\alpha, \beta \in B^\times$:

$$\begin{array}{ccc} I & \xrightarrow{\operatorname{nrd}_I} & L \\ {\scriptstyle \wr}\big\downarrow {\scriptstyle \alpha\cdot\beta} & & {\scriptstyle \wr}\big\downarrow {\scriptstyle ab} \\ J = \alpha I \beta & \xrightarrow{\operatorname{nrd}_J} & abL \end{array} \qquad (19.6.3)$$

19.6.4. Suppose that $\operatorname{nrd}(I) = L = aR$ is principal. Then there is a similarity

$$\begin{array}{ccc} I & \xrightarrow{\operatorname{nrd}_I} & L = aR \\ \big\| & & {\scriptstyle \wr}\big\downarrow {\scriptstyle a^{-1}} \\ I & \xrightarrow{a^{-1}\operatorname{nrd}_I} & R \end{array} \qquad (19.6.5)$$

In other words, if every value of the quadratic form is divisible by a, then up to similarity it is equivalent to consider the quadratic form $a^{-1}\operatorname{nrd}$, taking values in R.

Lemma 19.6.6. *Suppose I is invertible. Then the quadratic form* $\operatorname{nrd}_I : I \to L$ *is locally oriented similar to* $\operatorname{nrd}_{\mathcal{O}} : \mathcal{O} \to R$*, where* $\mathcal{O} = \mathcal{O}_{\mathsf{R}}(I)$.

Proof. By 19.6.3, if $I = \alpha\mathcal{O}$ is principal, then nrd_I is similar to $\operatorname{nrd}_{\mathcal{O}}$. If I is invertible, then I is locally principal, so for all primes $\mathfrak{p}$ of R the quadratic form $\operatorname{nrd}\colon I_{\mathfrak{p}} \to L_{\mathfrak{p}}$ is similar to $\operatorname{nrd}\colon \mathcal{O}_{\mathfrak{p}} \to R_{\mathfrak{p}}$ where $\mathcal{O}_{\mathfrak{p}}$ is the left (or right) order of $I_{\mathfrak{p}}$. (A similar statement holds on the left.) $\square$

19.6.7. From Lemma 15.3.6, it follows from Lemma 19.6.6 that

$$\operatorname{disc}(\operatorname{nrd}_I) = \operatorname{disc}(O)$$

and in particular this discriminant is a square.

The quadratic forms nrd_I are all locally similar, respecting the canonical orientation 5.6.7 on B. Therefore, the map

$$\begin{aligned} \operatorname{BrtCl} O &\to \left\{ \begin{array}{c} \text{Quaternary quadratic forms over } \mathbb{Z} \\ \text{locally similar to } \operatorname{nrd}_O \\ \text{up to oriented similarity} \end{array} \right\} \\ [I] &\mapsto \operatorname{nrd}_I \end{aligned}$$

is (well-defined and) surjective.

Remark 19.6.8. The Brandt groupoid is *connected* as a groupoid. This can also be viewed in the language of quadratic forms: a connected class of orders is equivalently a genus of integral ternary quadratic forms, and this is akin to a *resolvent* for the quaternary norm forms. We refer to Chapter 23 for further discussion.

Exercises

1. Verify the computational details in the example of section 19.2.
2. Let $B = (-1, -11 \mid \mathbb{Q})$ with $\operatorname{disc} B = 11$ and $O = \mathbb{Z}\langle i, (1 + j)/2\rangle$ a maximal order. Compute $\operatorname{BrtCl} O$, in a manner analogous to the example of section 19.2.
3. Let G be a groupoid.
   (a) Show that if $a, b, c \in G$ and both $a * b$ and $a * c$ are defined, then $b * b^{-1} = c * c^{-1}$ (and both are defined).
   (b) Show that for all $a \in G$ we have $(a^{-1})^{-1}$.
4. Let G be a group acting on a nonempty set X. Let
   $$A(G, X) = \{(g, x) : g \in G, x \in X\}.$$
   Show that $A(G, X)$ has a natural groupoid structure with $(g, x) * (h, y) = (gh, y)$ defined if and only if $x = hy$. What are the identity elements?
5. Show that in a homomorphism $\phi\colon G \to G'$ of groupoids, the set of identity elements of G maps to the set of identity elements of G'.
6. Let C be a small category. Show that there is a unique maximal subcategory that is a groupoid. *[Hint: Discard all nonisomorphisms.]*

7. Let X be a set and let $\sim$ be an equivalence relation on X, thought of as a subset $S \subseteq X \times X$. Equip S with the partial binary operation $*$ defined by $(x, y) * (y, z) = (x, z)$ for $(x, y), (y, z) \in S$ (and $(x, y) * (w, z)$ is not defined if $y \neq w$). Show that S is a groupoid. [This shows that "equivalence relations are groupoids", cf. (19.3.13).]
8. Let F be a field and let $\mathrm{GL}(F) = \bigcup_{n=1}^{\infty} \mathrm{GL}_n(F)$. Show that $\mathrm{GL}(F)$ has a natural structure of groupoid, sometimes called the **general linear groupoid** over F.
9. Show that the reduced norm is a homomorphism from the groupoid of invertible R-lattices in B to the group(oid) of fractional R-ideals in F.
10. Let X be a nonempty topological space, and let $x, y \in X$. Recall that a **path** from x to y is a continuous map $\upsilon_0 \colon [0,1] \to X$ with $\upsilon(0) = x$ and $\upsilon(1) = y$. We say that paths $\upsilon_0, \upsilon_1 \colon [0,1] \to X$ from x to y are **homotopic** if there exists a continuous map $H \colon [0,1] \times [0,1] \to X$ such that $H(0, s) = x$ and $H(1, s) = y$ for all $s \in [0,1]$ and $H(t, 0) = \upsilon_0$ and $H(t, 1) = \upsilon_1(t)$ for all $t \in [0,1]$. [So each $H(t, s)$ for fixed $t \in [0,1]$ is a path from x to y, and this set of paths varies continuously.]

    (a) Check that being homotopic defines an equivalence relation on the set of continuous paths from x to y.
    (b) Check that paths can be composed (going at twice speed) and that composition of paths is well-defined on homotopy classes.
    (c) Show that composition of homotopy classes of continuous paths is associative.

    Let $\Pi(X)$ be the category whose objects are the points of X and with morphisms to be the set of homotopy classes of continuous paths from x to y under composition.

    (d) Show that $\Pi(X)$ is a category.
    (e) Show that $\Pi(X)$ is a groupoid, called the **fundamental groupoid** of X.
    (f) Finally, for all $x \in X$, show the set of all morphisms from x to x in $\Pi(X)$ is a group (the more familiar **fundamental group** $\pi_1(X, x)$ with base point x).

11. Continuing the previous exercise, show that if X is path-connected, then $\Pi(X)$ is equivalent as a category to a groupoid with one object. *[Hint: choose a point $x \in X$, look at the group(oid) $\pi_1(X, x)$.]*

This chapter is licensed under the terms of the Creative Commons Attribution-NonCommercial 4.0 International License (http://creativecommons.org/licenses/by-nc/4.0/), which permits any noncommercial use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license and indicate if changes were made.

The images or other third party material in this chapter are included in the chapter's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the chapter's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.

Chapter 20

Integral representation theory

In this chapter, we consider a slightly more general framework on the preceding chapters: we consider lattices as projective modules, and relate this to invertibility and representation theory in an integral sense.

20.1 ▹ Projectivity, invertibility, and representation theory

Let R be a Dedekind domain with field of fractions $F = \operatorname{Frac} R$. Finitely generated, projective R-modules have played an important role throughout this text, and we now seek to understand them in the context of orders.

To this end, let B be a finite-dimensional F-algebra and let $O \subseteq B$ be an R-order. A **left O-lattice** M is an R-lattice that is a left O-module, i.e., M is a finitely generated, projective (locally free) R-module that has the structure of a left O-module. We make a similar definition on the right.

We say that a left (or right) O-lattice M is **projective** if it is a direct summand of a free left (or right) O-module. Projectivity for lattices in B is related to invertibility as follows (Theorem 20.3.3).

Theorem 20.1.1. *Let $I \subseteq B$ be an R-lattice. Then I is invertible if and only if I is projective as a left $O_{\mathsf{L}}(I)$-module and as a right $O_{\mathsf{R}}(I)$-module.*

One can also tease apart left and right invertibility if desired; in the quaternion context, these are equivalent anyway because of the standard involution (Main Theorem 20.3.9).

Given our efforts to understand invertible lattices, one may think that Theorem 20.1.1 is all there is to say. However, two issues remain. First, there may be finitely generated (projective) O-modules that are not lattices, and they play a structurally important role for the order O. Second, and this point is subtle: there may be lattices $I \subseteq B$ that are projective as a left O-module, but with $O_{\mathsf{L}}(I) \supsetneq O$; in other words, such lattices are invertible over a larger order, even though they still have good properties as modules over the smaller order.

© The Author(s) 2021

J. Voight, *Quaternion Algebras*, Graduate Texts in Mathematics 288,
https://doi.org/10.1007/978-3-030-56694-4_20

Example 20.1.2. Let

$$O := \begin{pmatrix} \mathbb{Z} & \mathbb{Z} \\ p\mathbb{Z} & \mathbb{Z} \end{pmatrix} \subseteq B := \mathrm{M}_2(\mathbb{Q})$$

be the order consisting of integral matrices that are upper triangular modulo a prime p. We will exhibit both of the issues above. First, we consider O as a left O-module: it decomposes as

$$\begin{aligned} O = O\begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix} \oplus O\begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix} &= \begin{pmatrix} \mathbb{Z} & 0 \\ p\mathbb{Z} & 0 \end{pmatrix} \oplus \begin{pmatrix} 0 & \mathbb{Z} \\ 0 & \mathbb{Z} \end{pmatrix} \\ &\simeq \begin{pmatrix} \mathbb{Z} \\ (p) \end{pmatrix} \oplus \begin{pmatrix} \mathbb{Z} \\ \mathbb{Z} \end{pmatrix} =: I_1 \oplus I_2. \end{aligned} \tag{20.1.3}$$

The two left O-modules I_1, I_2 are visibly projective, and they are not isomorphic: intuitively, an isomorphism would have to be multiplication on the left by a 2×2-matrix that commutes with multiplication O, and so it must be scalar. More precisely, suppose $\phi \in \mathrm{Hom}_O(I_1, I_2)$ is an isomorphism of left O-modules. Extending scalars, we have

$$\mathbb{Q}I_1 = \mathbb{Q}I_2 = \begin{pmatrix} \mathbb{Q} \\ \mathbb{Q} \end{pmatrix} =: V,$$

and the extension of ϕ gives an element in $\mathrm{Aut}_B(V)$ where $B = \mathrm{M}_2(\mathbb{Q}) = \mathrm{End}_{\mathbb{Q}}(V)$, so commutes with the action of B and is therefore central: which is to say ϕ is a scalar matrix, and that is absurd.

The lattice $I = \mathrm{M}_2(\mathbb{Z})$ is invertible as lattice, since it is an order (!); and it is a two-sided fractional O-ideal, but it is not sated. We claim that I is also a projective O-module: this follows from the fact that $\mathrm{M}_2(\mathbb{Z}) \simeq I_2^{\oplus 2}$ as a left O-module, so $\mathrm{M}_2(\mathbb{Z})$ is isomorphic to a direct summand of $O^{\oplus 2}$.

In this chapter, we establish some basic vocabulary of modules in the language of the representation theory of an order. In the case of algebras over a field, we defined a Jacobson radical as a way to measure the failure of the algebra to be semisimple. Similarly, for every ring A, we define the **Jacobson radical** $\operatorname{rad} A$ to be the intersection of all maximal left ideals of A: it again measures the failure of left indecomposable modules to be simple. There is a left-right symmetry to $\operatorname{rad} A$, and in fact $\operatorname{rad} A \subseteq A$ is a two-sided A-ideal.

Locally, the Jacobson radical plays a key role. Suppose R is a complete DVR with unique maximal ideal $\mathfrak{p}$. Then $\mathfrak{p} = \operatorname{rad} O$ since it is the maximal ideal. Moreover, we will see that $\mathfrak{p}O \subseteq \operatorname{rad} O$, so $O/\operatorname{rad} O$ is a finite-dimensional semisimple k-algebra. Much of the structure of O-modules is reflected in the structure of modules over the quotient $O/\operatorname{rad} O$ (see Lemma 20.6.8).

Remark 20.1.4. In representation theory, generally speaking, to study the action of a group on some kind of object (vector space, simplicial complex, etc.) one introduces some kind of group ring and studies modules over this ring. The major task becomes to classify such modules. For example, let R be a Dedekind domain with $F = \operatorname{Frac} R$, and let O be an R-order in a finite-dimensional F-algebra B. A finitely generated **integral representation** of O is a finitely generated O-module that is projective as

an R-module (in particular, is R-torsion free). The integral representations of O are quite complicated! Nevertheless, integral representation theory is a beautiful blend of number theory, commutative algebra, and linear algebra. In section 20.6, we will see some of the basic ingredients when R is a DVR, and in section 21.4 in the next chapter we will show that *hereditary orders* have a tidy integral representation theory. For more on the subject, see the surveys by Reiner [Rei70, Rei76] as well as the massive treatises by Curtis–Reiner [CR81, CR87].

20.2 Projective modules

As we will need the notion over several different rings, we start more generally: let A be a ring (not necessarily commutative, but with 1). For an introduction to the theory of projective modules and related subjects, see Lam [Lam99, §2] and Curtis–Reiner [CR81, §2], and Berrick–Keating [BK2000, §2].

Definition 20.2.1. Let P be a finitely generated left A-module. Then P is **projective** as a left A-module if it is a direct summand of a free left A-module.

A finitely generated free module is projective. The notion of projectivity is quite fundamental, as the following proposition indicates.

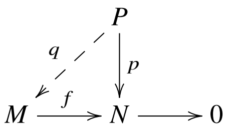
Proposition 20.2.2. *Let P be a finitely generated left A-module. Then the following are equivalent:*

(i) *P is projective;*

(ii) *There exists a finitely generated left A-module Q such that $P \oplus Q$ is free as a left A-module.*

(iii) *Every surjective homomorphism $f\colon M \to P$ (of left A-modules) has a splitting $g : P \to M$ (i.e., $f \circ g = \mathrm{id}_P$);*

(iv) *Every diagram*



of left A-modules with exact bottom row can be extended as indicated, with $p = f \circ q$; and

(v) $\mathrm{Hom}_A(P, -)$ *is a (right) exact functor.*

Proof. See Lam [Lam99, Chapter 2]. In statement (v), given a short exact sequence

$$0 \to Q \to M \to N \to 0$$

then $\mathrm{Hom}_A(P, -)$ is always left exact, so

$$0 \to \mathrm{Hom}_A(P, Q) \to \mathrm{Hom}_A(P, M) \to \mathrm{Hom}_A(P, N) \qquad (20.2.3)$$

is exact; the condition for P to be projective is that $\operatorname{Hom}_A(P,-)$ is right exact, so the full sequence

$$0 \to \operatorname{Hom}_A(P,Q) \to \operatorname{Hom}_A(P,M) \to \operatorname{Hom}_A(P,N) \to 0 \tag{20.2.4}$$

is short exact. □

20.2.5. A finite direct sum $P = \bigoplus_i P_i$ of finitely generated A-modules is projective if and only if each summand P_i is projective: indeed, the functor $\operatorname{Hom}_R(P,-)$ is naturally isomorphic to $\prod_i \operatorname{Hom}_R(P_i,-)$, so we apply condition (v) of Proposition 20.2.2.

20.2.6. Localizing Proposition 20.2.2(v), and using the fact that a sequence is exact if and only if it is exact locally (Exercise 20.1(a)), we see that P is projective as a left O-module if and only if $P_{(\mathfrak{p})}$ is projective as a left $O_{(\mathfrak{p})}$-module for all primes $\mathfrak{p} \subseteq R$

Definition 20.2.7. A **left O-lattice** is an R-lattice M that is a left O-module.

We make a similar definition on the right.

20.2.8. A left O-lattice M is **locally free** of rank $r \geq 1$ if $M_{\mathfrak{p}} \simeq O_{\mathfrak{p}}^{\oplus r}$ as left O-modules for all primes $\mathfrak{p} \subseteq R$. If follows from 20.2.5 and 20.2.6 that a locally free O-lattice is projective.

20.3 Projective modules and invertible lattices

Now let R be a noetherian domain with $F := \operatorname{Frac} R$, let B be a finite-dimensional F-algebra, and let $O \subseteq B$ be an R-order.

One can extend the base ring of the module while preserving projectivity, as follows.

Lemma 20.3.1. *Let $O \subseteq O'$ be R-orders in B and let M be a left O'-lattice. If M is projective as a left O-module, then M is projective as a left O'-module.*

Proof. Suppose M is projective as a left O-module; then $M \oplus N \simeq O^r$ for some $r \geq 0$. Tensor with O' to get

$$(O' \otimes_O M) \oplus (O' \otimes_O N) \simeq (O')^r. \tag{20.3.2}$$

Since multiplication gives an isomorphism of left O'-modules $O' \otimes_O M \xrightarrow{\sim} O'M = M$, the result follows. (More generally, see Harada [Har63a, Lemma 1.3].) □

In the commutative case, an R-lattice $\mathfrak{a} \subseteq F$ is invertible as an R-module if and only if $\mathfrak{a}$ is projective as a (left and right) R-module. Something is true in this more general context.

Theorem 20.3.3. *Let $I \subseteq B$ be an R-lattice.*

(a) *$I^{-1}I = O_{\mathsf{R}}(I)$ if and only if I is projective as a left $O_{\mathsf{L}}(I)$-module, and $II^{-1} = O_{\mathsf{L}}(I)$ if and only if I is projective as a right $O_{\mathsf{R}}(I)$-module.*

(b) *I is projective as a left $O_{\mathsf{L}}(I)$-module and a right $O_{\mathsf{R}}(I)$-module if and only if I is invertible (as an R-lattice).*

The difference between (a) and (b) in Theorem 20.3.3 is the *compatibility* of the two products.

Proof. We begin with (a). To prove the implication ($\Rightarrow$), suppose $I^{-1}I = O_{\mathsf{R}}(I)$; then there exist $\alpha_i \in I$ and $\alpha_i^* \in I^{-1}$ such that $\sum_i \alpha_i^*\alpha_i = 1$. We may extend the set α_i to generate I as a left $O_{\mathsf{L}}(I)$-module by taking $\alpha_i^* = 0$ if necessary. We define the surjective map

$$\begin{aligned} f\colon M = \bigoplus_i O_{\mathsf{L}}(I)e_i &\to I \\ e_i &\mapsto \alpha_i. \end{aligned} \tag{20.3.4}$$

Consider the map

$$\begin{aligned} g : I &\to M \\ \beta &\mapsto \sum_i \beta\alpha_i^* e_i; \end{aligned}$$

the map g is defined because for all $\beta \in I$, we have $\beta\alpha_i^* \in II^{-1}$, and as always $II^{-1}I \subseteq I$ so $II^{-1} \subseteq O_{\mathsf{L}}(I)$. The map g is a splitting of f since

$$(f \circ g)(\beta) = \sum_i \beta\alpha_i^*\alpha_i = \beta \sum_i \alpha_i^*\alpha_i = \beta.$$

Therefore I is a direct summand of M, so I is projective as a left $O_{\mathsf{L}}(I)$-module.

Next we prove ($\Leftarrow$). There exists a nonzero $r \in I \cap R$ (Exercise 9.2), so to show that I is left invertible, we may replace I with $r^{-1}I$ and therefore suppose that $1 \in I$. Following similar lines as above, let α_i generate I as a left $O_{\mathsf{L}}(I)$-module, and consider the surjective map $f\colon M = \bigoplus_i O_{\mathsf{L}}(I)e_i \to I$ by $e_i \mapsto \alpha_i$. Then since I is projective as a left $O_{\mathsf{L}}(I)$-module, this map splits by a map $g : I \to M$; suppose that $g(1) = (\alpha_i^*)_i$ with $\alpha_i^* \in O_{\mathsf{L}}(I)$; then

$$(f \circ g)(1) = 1 = \sum_i \alpha_i^*\alpha_i. \tag{20.3.5}$$

For all $\beta \in I$, we have $g(\beta) = (\beta\alpha_i^*)_i \in M$, so $\beta\alpha_i^* \in O_{\mathsf{L}}(I)$ for all i; therefore for all $\alpha, \beta \in I$ we have $\beta\alpha_i^*\alpha \in O_{\mathsf{L}}(I)I \subseteq I$, whence $\alpha_i^* \in I^{-1}$ by definition. Thus from (20.3.5) we have $1 \in I^{-1}I$, whence

$$O_{\mathsf{R}}(I) \subseteq I^{-1}IO_{\mathsf{R}}(I) = I^{-1}I \subseteq O_{\mathsf{R}}(I)$$

and thus equality holds.

For part (b), the implication ($\Leftarrow$) follows from (a), and the implication ($\Rightarrow$) for compatibility follows from Proposition 16.5.8. $\square$

Remark 20.3.6. The proof of Theorem 20.3.3 follows what is sometimes called the **dual basis lemma** for a projective module: see Lam [Lam99, (2.9)], Curtis–Reiner [CR81, (3.46)], or Faddeev [Fad65, Proposition 18.2].

20.3.7. Let $O, O' \subseteq B$ be R-orders. A O, O'-**bimodule** over R is an abelian group M with a left O-module and a right O'-module structure with the same action by R on the left and right (i.e., acting centrally, so $rm = mr$ for all $r \in R$ and $m \in M$). The R-lattice $I \subseteq B$ is an $O_{\mathsf{L}}(I), O_{\mathsf{R}}(I)$-bimodule over R.

When the equivalent conditions of Theorem 20.3.3(b) hold, we say that I is **projective** as a $O_{\mathsf{L}}(I), O_{\mathsf{R}}(I)$-bimodule over R.

Remark 20.3.8. In Theorem 20.3.3, we only considered an R-lattice I as a module over its left and right orders (i.e., we considered only *sated* fractional O, O'-ideals), for the reasons explained in 16.5.18.

Although invertible requires working in this way, it is possible for an R-lattice I to be projective as a left O-module but still $O \subsetneq O_{\mathsf{L}}(I)$: for example, if O is a hereditary order (see Chapter 21) contained properly in a maximal order $O \subsetneq O'$, then O' is projective as a left O-module.

Although this may seem a bit complicated, it is refreshing that for quaternion algebras, all of the sided notions coincide. We recall the equivalences in Main Theorem 16.7.7), building upon them.

Main Theorem 20.3.9. *Suppose R is a Dedekind domain and B is a quaternion algebra over $F = \operatorname{Frac} R$, and let $I \subset B$ be an R-lattice. Then the following are equivalent:*

- (ii) *I is invertible;*
- (iii) *I is left invertible;*
- (iii′) *I is right invertible;*
- (v) *I is projective as a left $O_{\mathsf{L}}(I)$-module; and*
- (v′) *I is projective as a right $O_{\mathsf{R}}(I)$-module.*

Proof. The equivalences (ii) ⇔ (ii′) ⇔ (iii) are from Main Theorem 16.7.7 (proven in Lemma 16.7.5). Theorem 20.3.3(a) gives (v) ⇒ (iii) and (v′) ⇒ (iii′), and Theorem 20.3.3(b) gives (ii) ⇒ (v), (v′). □

Example 20.3.10. Consider again Example 16.5.12. The lattice I has $O_{\mathsf{L}}(I) = O_{\mathsf{R}}(I) = O$ (so has the structure of a sated O, O-bimodule) but I is *not* invertible; from Main Theorem 20.3.9, it follows that I is *not* projective as a left or right O-module.

20.4 Jacobson radical

Before proceeding further in our analysis of orders, we pause to extend some notions in sections 7.2 and 7.4 from algebras to rings. We follow Reiner [Rei2003, §6a]; see also Curtis–Reiner [CR81, §5].

Throughout, let A be a ring (not necessarily commutative, but with 1).

Definition 20.4.1. Let M be a left A-module. We say M is **irreducible** or **simple** if $M \neq \{0\}$ and M contains no A-submodules except $\{0\}$ and M. We say M is **indecomposable** if whenever $M = M_1 \oplus M_2$ with M_1, M_2 left A-modules, then either $M_1 = \{0\}$ or $M_2 = \{0\}$.

20.4.2. We generalize Lemma 7.2.7. If I is a maximal left ideal of A, then A/I is a simple A-module. Conversely, if M is a simple A-module, then for any $x \in M$ nonzero we have $Ax = M$; therefore $M \simeq A/I$ where

$$I = \operatorname{ann}(x) := \{\alpha \in A : \alpha x = 0\}.$$

Definition 20.4.3. The **Jacobson radical** $\operatorname{rad} A$ is the intersection of all maximal left ideals of A. The ring A is **Jacobson semisimple** if $\operatorname{rad} A = \{0\}$.

Lemma 20.4.4. *The Jacobson radical* $\operatorname{rad} A$ *is the intersection of all annihilators of simple left A-modules;* $\operatorname{rad} A \subseteq A$ *is a two-sided A-ideal.*

Proof. The same proof as in Lemma 7.4.5 and Corollary 7.4.6 applies, *mutatis mutandis*. ◻

Example 20.4.5. If A is a commutative local ring, then $\operatorname{rad} A$ is the unique maximal ideal of A.

Example 20.4.6. Let R be a complete DVR with maximal ideal $\mathfrak{p} = \operatorname{rad} R$. Let $F = \operatorname{Frac} R$ and let D be a division algebra over F. Let $O \subseteq D$ be the valuation ring, the unique maximal R-order (Proposition 13.3.4). Then O has a unique two-sided ideal P by 13.3.10, and so $\operatorname{rad} O = P$.

Lemma 20.4.7. $A/\operatorname{rad} A$ *is Jacobson semisimple.*

Proof. Let $J = \operatorname{rad} A$. Since $JM = \{0\}$ for each simple left A-module M, we may view each such M as a simple left A/J-module. Now let $\alpha \in A$ be such that $\alpha + J \in \operatorname{rad}(A/J)$; then $(\alpha + J)M = \{0\}$, so $\alpha M = \{0\}$ and $\alpha \in J$; thus $\operatorname{rad}(A/J) = \{0\}$, and A/J is Jacobson semisimple. ◻

Lemma 20.4.8. *We have*

$$\operatorname{rad} A = \{\beta \in A : 1 - \alpha_1\beta\alpha_2 \in A^\times \textit{ for all } \alpha_1, \alpha_2 \in A\}.$$

Proof. See Exercise 20.6. ◻

Corollary 20.4.9. $\operatorname{rad} A$ *is the intersection of all maximal right ideals of A.*

Proof. Lemma 20.4.8 gives a left-right symmetric characterization of $\operatorname{rad} A$. ◻

Corollary 20.4.10. *If $\phi\colon A \to A'$ is a surjective ring homomorphism, then* $\phi(\operatorname{rad} A) \subseteq \operatorname{rad} A'$ *and we have an induced surjective homomorphism* $A/\operatorname{rad} A \to A'/\operatorname{rad} A'$.

Proof. Let $\beta \in \operatorname{rad} A$, let $\alpha_1', \alpha_2' \in A'$; since ϕ is surjective, there exist preimages $\alpha_1, \alpha_2 \in A$. By Lemma 20.4.8, $1 - \alpha_1\beta\alpha_2 \in A^\times$ and

$$\phi(1 - \alpha_1\beta_1\alpha_2) = 1 - \alpha_1'\phi(\beta)\alpha_2' \in A'^\times,$$

so by the same lemma, $\phi(\beta) \in \operatorname{rad} A'$. ◻

Corollary 20.4.11. *Let $I \subseteq A$ be a two-sided A-ideal.*

(a) *If A/I is Jacobson semisimple, then* $\operatorname{rad} A \subseteq I$.
(b) *If* $I \subseteq \operatorname{rad} A$, *then* $(\operatorname{rad} A)/I = \operatorname{rad}(A/I)$.

Proof. We have a surjection $\phi\colon A \to A/I$. For (a), we get $\phi(\operatorname{rad} A) \subseteq \operatorname{rad}(A/I) = \{0\}$ from Corollary 20.4.10, so $\operatorname{rad} A \subseteq I$. For (b), we get $\operatorname{rad}(A)/I \subseteq \operatorname{rad}(A/I)$ from the surjection, and applying (a) to $(A/I)/(\operatorname{rad}(A)/I)$ we get $\operatorname{rad}(A/I) \subseteq \operatorname{rad}(A)/I$. □

Lemma 20.4.12. (Nakayama's lemma). *Let M be a finitely generated left A-module such that* $(\operatorname{rad} A)M = M$. *Then* $M = \{0\}$.

Proof. If $M \neq \{0\}$, let $x_1, \dots, x_n$ be a minimal set of generators for M as a left A-module. Since $x_1 \in M = (\operatorname{rad} A)M$, we may write

$$x_1 = \beta_1 x_1 + \cdots + \beta_n x_n$$

with $\beta_i \in \operatorname{rad} A$. But then $1 - \beta_1 \in A^\times$, so the generator x_1 is redundant, a contradiction. □

Corollary 20.4.13. *Let M be a finitely generated left A-module, and let $N \subseteq M$ be a submodule such that* $N + (\operatorname{rad} A)M = M$. *Then* $N = M$.

Proof. By hypothesis, M/N is finitely generated, and $(\operatorname{rad} A)(M/N) = M/N$, so by Nakayama's lemma, $M/N = \{0\}$ and $M = N$. □

Lemma 20.4.14. *Let I be a maximal two-sided ideal of A. Then I contains* $\operatorname{rad} A$.

Proof. If I does not contain $\operatorname{rad} A$, then $I + \operatorname{rad} A$ is a two-sided ideal of A containing $\operatorname{rad} A$ and properly containing I. Since I is maximal, we have $I + \operatorname{rad} A = A$. By (the corollary to) Nakayama's lemma, we get $I = A$, a contradiction. □

20.5 Local Jacobson radical

Suppose now that R is a complete DVR with fraction field $F = \operatorname{Frac} R$, maximal ideal $\mathfrak{p} = \operatorname{rad} R$, and residue field $k = R/\mathfrak{p}$. Let B be a finite-dimensional F-algebra, and let $O \subseteq B$ be an R-order.

In this setting, we may identify the Jacobson radical via pullback as follows.

Theorem 20.5.1. *Let $\phi\colon O \to O/\mathfrak{p}O$ be reduction modulo $\mathfrak{p}$. Then*

$$\operatorname{rad} O = \phi^{-1}(\operatorname{rad} O/\mathfrak{p}O) \supseteq \mathfrak{p}O,$$

and $(\operatorname{rad} O)^r \subseteq \mathfrak{p}O$ *for some* $r > 0$.

Proof. See Reiner [Rei2003, Theorem 6.15]. □

Corollary 20.5.2. $O/\operatorname{rad} O$ *is a (finite-dimensional) semisimple k-algebra.*

Proof. Since $\operatorname{rad} O \supseteq \mathfrak{p}O$, we conclude that $O/\mathfrak{p}O$ is a k-algebra; it is Jacobson semisimple by 20.4.7 and hence semisimple by Lemma 7.4.2. □

Definition 20.5.3. A two-sided ideal $J \subseteq O$ is **topologically nilpotent** if $J^r \subseteq \mathfrak{p}O$ for some $r > 0$.

Remark 20.5.4. The order O as a free R-module has a natural topology induced from the $\mathfrak{p}$-adic topology on R; J is topologically nilpotent if and only if $J^r \to \{0\}$ in this topology.

Corollary 20.5.5. *Let $I \subseteq O$ be a two-sided ideal. Then the following are equivalent:*

(a) $I \subseteq \operatorname{rad} O$*;*
(b) $I^r \subseteq \operatorname{rad} O$ *for some* $r > 0$*; and*
(c) I *is topologically nilpotent.*

Proof. See Reiner [Rei2003, Exercise 39.1, Exercise 6.3]. □

20.6 Integral representation theory

We continue our notation that R is a complete DVR. We now turn to some notions in integral representation theory. In this local case, there is a tight connection between the representation theory of O (viewed in terms of O-modules) and the representation theory of the quotient $O/\mathfrak{p}O$ which is a k-algebra of finite dimension over k, since O is finitely generated as an R-module.

20.6.1. Recall that a representation of B over F is the same as a left B-module. If M is a finitely-generated left O-module, then $V := M \otimes_R F$ is a left B-module, and $M \subseteq V$ is an R-lattice. A O-**supermodule** of M is a left O-module $V \supseteq M' \supseteq M$.

The following result is foundational.

Theorem 20.6.2. (Krull–Schmidt). *Every finitely generated left O-module M is expressible as a finite direct sum of indecomposable modules, uniquely determined by M up to O-module isomorphism and reordering.*

Proof. Since M is finitely generated over R it is itself noetherian, so the process of decomposing M into direct summands terminates. See Curtis–Reiner [CR81, (6.12)] or Reiner [Rei2003, §6, Exercise 6] for hints that lead to a proof of the second (uniqueness) part. □

Corollary 20.6.3. *Let $M = M_1 \oplus \cdots \oplus M_r$ be a decomposition into finitely generated indecomposable left O-modules, and let $N \subseteq M$ be a direct summand. Then $N \simeq M_{i_1} \oplus \cdots \oplus M_{i_s}$ for some subset $\{i_1, \dots, i_s\} \subseteq \{1, \dots, n\}$.*

Proof. By hypothesis, we can write $\bigoplus_{i=1}^n M_i = N \oplus N'$, with N' a finitely generated left O-module. By the Krull–Schmidt theorem (Theorem 20.6.2), if we write N, N' as the direct sums of indecomposable modules, the conclusion follows. □

20.6.4. We saw in 7.2.19 that idempotents govern the decomposition of the F-algebra B into indecomposable left B-modules. The same argument shows that a decomposition

$$O = P_1 \oplus \cdots \oplus P_r \tag{20.6.5}$$

into a direct sum of indecomposable left O-modules corresponds to an idempotent decomposition $1 = e_1 + \cdots + e_r$, with the e_i a complete set of primitive orthogonal idempotents. Moreover, each $P_i = Oe_i$ is a projective indecomposable left O-module.

Conversely, if P is a projective indecomposable finitely generated left O-module, then $P \simeq P_i$ for some i: taking a set of generators we have a surjective O-module homomorphism $O^r \to P$, and since P is projective we have $P \subseteq O^r$ a direct summand, so Corollary 20.6.3 applies.

Consequently, if P is a projective left O-lattice, then $P \simeq P_1^{\oplus n_1} \oplus \cdots \oplus P_r^{\oplus n_r}$ with $n_i \geq 0$ for $i = 1, \dots, r$.

20.6.6. The decomposition of an order into projective indecomposables is a nice way to keep track of other orders, as follows. We extend our notation slightly, and define

$$O_{\mathsf{L}}(M) := \{\alpha \in B : \alpha M \subseteq M\}$$

for every left O-submodule $M \subseteq B$.

Take a decomposition of O in (20.6.5); since each P_i is a left O-module, extending scalars it is a left B-module, so

$$\bigcap_{i=1}^{r} O_{\mathsf{L}}(P_i) = O. \tag{20.6.7}$$

Now let $I \subseteq B$ be an R-lattice with $O \subset O_{\mathsf{L}}(I)$ that is projective as an O-module. By 20.6.4, considering I as a left O-module, we have an isomorphism of left O-modules

$$\phi \colon I \xrightarrow{\sim} P_1^{\oplus n_1} \oplus \cdots \oplus P_r^{n_r}$$

with $n_i \geq 0$. We claim that

$$O_{\mathsf{L}}(I) = \bigcap_{\substack{i \\ n_i > 0}} O_{\mathsf{L}}(P_i).$$

Indeed, we have $\alpha I \subseteq I$ if and only if $\phi(\alpha I) = \alpha\phi(I) \subseteq \phi(I)$, since ϕ is a O-module homomorphism so extends to a B-algebra homomorphism, and finally $\alpha\phi(I) \subseteq \phi(I)$ if and only if $\alpha P_i \subseteq P_i$ for all i with $n_i > 0$, as in (20.6.7).

We now relate a decomposition of O into a decomposition of $O/\mathfrak{p}O$.

Lemma 20.6.8. *Let $J = \operatorname{rad} O$. The association $I \mapsto I/JI$ gives a bijection between isomorphism classes of indecomposable finitely generated projective left O-modules and isomorphism classes of simple finite-dimensional left O/J-modules.*

Proof. The proof requires a bit of fiddling with idempotents, but is otherwise straightforward—so it makes a good exercise (Exercise 20.7). ∎

Corollary 20.6.9. *If I is projective indecomposable, then $JI \subsetneq I$ is the unique maximal O-submodule of I.*

Proof. By Lemma 20.6.8, since I is indecomposable, I/JI is simple so JI is a unique maximal submodule. If $I' \subsetneq I$ is another maximal O-submodule, then $JI + I' = I$, and by Nakayama's lemma $I' = I$, a contradiction. □

We finish our local study over R a complete DVR with composition series for modules over an order.

Definition 20.6.10. Let M be an O-lattice. A **composition series** for M is a strictly decreasing sequence

$$M = M_0 \supsetneq M_1 \supsetneq M_2 \supsetneq \dots$$

such that $\bigcap_{i=1}^{\infty} M_i = \{0\}$ and each **composition factor** M_i/M_{i+1} is simple as a O-module.

The **length** of a composition series is the largest integer r such that $M_r = \{0\}$ if r exists (in which case we call the series **finite**), and otherwise the length is ∞.

20.6.11. If M has a finite composition series, then its **length** $\ell(M)$ is well-defined, independent of the series. For example, taking $R = F$ and $O = B$, a finitely generated B-module is a finite-dimensional F-vector space, so every composition series is finite and every B-module V has a well-defined length $\ell(V)$.

20.6.12. Let $N \subsetneq M$ be a maximal O-submodule. We claim that $JM \subseteq N$. Otherwise, $N + JM = M$ by maximality, so by Nakayama's lemma (Corollary 20.4.13), $N = M$, a contradiction.

20.7 ∗ Stable class group and cancellation

To conclude this chapter, we apply the above results and consider a different way to form of a group of ideal classes; for further reference on the topics of this section, see Curtis–Reiner [CR87, §§49–51] or Reiner [Rei2003, §38].

Let R be a Dedekind domain with field of fractions F.

20.7.1. Recall that the group $\operatorname{Cl} R$ records classes of fractional ideals, or what is more relevant here, isomorphism classes of projective modules of rank 1. Here is another way to see the group law on $\operatorname{Cl} R$: given two such fractional ideals $\mathfrak{a}, \mathfrak{b}$ up to isomorphism, there is an isomorphism of R-modules

$$\mathfrak{a} \oplus \mathfrak{b} \simeq R \oplus \mathfrak{a}\mathfrak{b},$$

and the class of $\mathfrak{a}\mathfrak{b}$ is uniquely determined by this isomorphism by 9.3.10.

We now consider an analogous construction to 20.7.1 in the noncommutative setting. Let B be a simple F-algebra and $O \subseteq B$ an R-order. We begin with a technical lemma.

Lemma 20.7.2. (Weak approximation). *Let I be a locally principal left fractional O-ideal and let $\mathfrak{a} \subseteq R$ be an ideal. Then there exists $\beta \in B^\times$ such that $I\beta \subseteq O$ and*

$$(I\beta)_\mathfrak{p} = O_\mathfrak{p} \quad \textit{for all } \mathfrak{p} \mid \mathfrak{a}. \tag{20.7.3}$$

Proof. For each prime $\mathfrak{p}$, we have $I_\mathfrak{p} = O_\mathfrak{p}\alpha_\mathfrak{p}$ with $\alpha_\mathfrak{p} \in B_\mathfrak{p}^\times$. Because F is dense in $F_\mathfrak{p}$, there exists $\beta \in (O : I)_\mathsf{R}$ such that $\alpha_\mathfrak{p}\beta_\mathfrak{p} \equiv 1 \pmod{\mathfrak{p}O_\mathfrak{p}}$ for all $\mathfrak{p} \mid \mathfrak{a}$. By norms, we have $\beta \in B^\times$. Letting $\mu_\mathfrak{p} := \alpha_\mathfrak{p}\beta_\mathfrak{p}$, we have $\mu_\mathfrak{p} - 1 \in \mathfrak{p}O_\mathfrak{p} \supseteq \operatorname{rad} O_\mathfrak{p}$ by Theorem 20.5.1 so $\mu_\mathfrak{p} \in O_\mathfrak{p}^\times$ by Lemma 20.4.8. Therefore $(I\beta)_\mathfrak{p} = O_\mathfrak{p}$ for all $\mathfrak{p} \mid \mathfrak{a}$. □

Proposition 20.7.4. *If $I, I' \subseteq B$ are locally principal left fractional O-ideals, then there exists a locally principal left fractional O-ideal J and an isomorphism*

$$I \oplus I' \simeq J \oplus O \tag{20.7.5}$$

of left O-modules.

Proof. We may suppose without loss of generality that $I, I' \subseteq O$. Then we have exact sequences of left O-modules

$$0 \to I \xrightarrow{\phi} O \to O/I \to 0$$

$$0 \to I' \xrightarrow{\phi'} O \to O/I' \to 0$$

The module O/I is R-torsion, annihilated by the (nonzero) R-ideal $\mathfrak{a} = [O : I]_R$, and similarly with I', annihilated by $\mathfrak{a}' = [O : I']_R$. By weak approximation (Lemma 20.7.2), replacing I' with $I'\beta$ we may suppose that $I'_\mathfrak{p} = O_\mathfrak{p}$ for all $\mathfrak{p} \mid \mathfrak{a}$, and hence $\mathfrak{a}, \mathfrak{a}'$ are coprime. Then for all primes $\mathfrak{p}$ of R, we have either $(O/I)_\mathfrak{p} = \{0\}$ so $\phi_\mathfrak{p}$ is surjective, or correspondingly $\phi'_\mathfrak{p}$ is surjective.

Now consider the left O-module homomorphism

$$\phi + \phi' : I \oplus I' \to O \tag{20.7.6}$$

obtained by summing the two natural inclusions. We just showed that $(\phi + \phi')_\mathfrak{p}$ is surjective for all primes $\mathfrak{p}$, so it follows that $\phi + \phi'$ is surjective: the cokernel $M := \operatorname{coker}(\phi + \phi')$ has $M_\mathfrak{p} = \{0\}$ for all $\mathfrak{p}$, and $M = \{0\}$. Moreover, since O is projective as a left O-module, the map $\phi + \phi'$ splits (or note that the map splits locally for every prime $\mathfrak{p}$, so it splits globally, Exercise 20.1(b)). If we let $J := \ker(\phi + \phi')$, we then obtain an isomorphism

$$I \oplus I' \simeq J \oplus O. \tag{20.7.7}$$

To conclude, we show that J is locally principal. To this end, we localize at a prime $\mathfrak{p}$ and note that I, I' are locally principal, so

$$I_\mathfrak{p} \oplus I'_\mathfrak{p} \simeq O_\mathfrak{p}^{\oplus 2} \simeq J_\mathfrak{p} \oplus O_\mathfrak{p}. \tag{20.7.8}$$

But by the Krull–Schmidt theorem (Theorem 20.6.2) and Exercise 20.8, we can cancel one copy of $O_\mathfrak{p}$ from both sides! We conclude that $J_\mathfrak{p} \simeq O_\mathfrak{p}$ as left O-modules and therefore by Lemma 17.3.3 that $J_\mathfrak{p}$ is (right) principal. □

The candidate binary operation in Proposition 20.7.4 has a simple description in the "coprime" case.

Lemma 20.7.9. *Let $I, I' \subseteq O$ be locally principal integral left O-ideals, and suppose for every prime $\mathfrak{p} \subseteq R$ either $I_{\mathfrak{p}} = O_{\mathfrak{p}}$ or $I'_{\mathfrak{p}} = O_{\mathfrak{p}}$. Then*

$$I \oplus I' \simeq O \oplus J, \quad \text{where } J = I \cap I'.$$

Moreover, writing $I_{\mathfrak{p}} = O\alpha_{\mathfrak{p}}$ and $I'_{\mathfrak{p}} = O\alpha'_{\mathfrak{p}}$, we have

$$J_{\mathfrak{p}} = O_{\mathfrak{p}}\alpha_{\mathfrak{p}}\alpha'_{\mathfrak{p}} = O_{\mathfrak{p}}\alpha'_{\mathfrak{p}}\alpha_{\mathfrak{p}}.$$

By weak approximation (Lemma 20.7.2), the hypothesis of Lemma 20.7.9 can always be arranged to hold for I, I', up to isomorphism (as left O-ideals).

Proof. By hypothesis, if $\phi, \phi \colon I, I' \hookrightarrow O$ are the inclusions, then the map $\phi + \phi' \colon I \oplus I' \to O$ as in (20.7.6) is surjective. We have

$$\ker(\phi + \phi') = \{(\alpha, \alpha') \in I \oplus I' : \alpha + \alpha' = 0\} \simeq I \cap I'$$

by projection onto either coordinate, since $\alpha = -\alpha' \in I \cap I'$. This gives an exact sequence

$$0 \to I \cap I' \to I \oplus I' \to O \to 0$$

and as above $I \oplus I' \simeq J \oplus O$ with $J = I \cap I'$. The final statement follows from the hypothesis that either $I_{\mathfrak{p}} = O_{\mathfrak{p}}$ or $I'_{\mathfrak{p}} = O_{\mathfrak{p}}$, since then $\alpha_{\mathfrak{p}} \in O_{\mathfrak{p}}^{\times}$ or $\alpha'_{\mathfrak{p}} \in O_{\mathfrak{p}}^{\times}$. ◻

In order to get a well-defined binary operation, we need an equivalence relation: we will need to identify J, J' if $J \oplus O \simeq J' \oplus O$. But the copies of O needed for the axioms start to pile up, so we make the following more general definition.

Definition 20.7.10. Let $J, J' \subseteq B$ be locally principal left O-ideals. We say that J is **stably isomorphic** to J' if there exists an isomorphism of left O-modules

$$J \oplus O^{\oplus r} \simeq J' \oplus O^{\oplus r}$$

for some $r \geq 0$.

Let $[J]_{\mathsf{St}}$ denote the stable isomorphism class of a left O-ideal J and let $\operatorname{StCl} O$ be the set of stable isomorphism classes of left O-ideals in B.

Proposition 20.7.11. $\operatorname{StCl} O$ *is an abelian group under the binary operation* (20.7.5)*, written* $[I]_{\mathsf{St}} + [I']_{\mathsf{St}} = [J]_{\mathsf{St}}$*, with identity* $[O]_{\mathsf{St}}$.

Accordingly, we call $\operatorname{StCl} O$ the **stable class group** of O; it is also referred to as the **locally free class group** of O.

Proof. The operation is well-defined: if $[I_1]_{\mathsf{St}} = [I_2]_{\mathsf{St}}$ via $I_1 \oplus O^{\oplus r} \simeq I_2 \oplus O^{\oplus r}$ and the same with $[I_1']_{\mathsf{St}} = [I_2']_{\mathsf{St}}$, and we perform the binary operation $I_1 \oplus I_1' \simeq J_1 \oplus O$ and the same with the subscripts $_2$, then

$$\begin{aligned} J_1 \oplus O^{\oplus(r+r'+1)} &\simeq (I_1 \oplus O^{\oplus r}) \oplus (I_1' \oplus O^{\oplus r'}) \\ &\simeq (I_2 \oplus O^{\oplus r}) \oplus (I_2' \oplus O^{\oplus r'}) \\ &\simeq J_2 \oplus O^{\oplus(r+r'+1)} \end{aligned} \tag{20.7.12}$$

so $[J_1]_{\mathsf{St}} = [J_2]_{\mathsf{St}}$. It is similarly straightforward to verify that the operation is associative and commutative and that $[O]_{\mathsf{St}}$ is the identity.

To conclude, we show that $\operatorname{StCl} O$ has inverses. Let $I \subseteq O$ be a locally principal O-ideal. For each prime $\mathfrak{p} \subseteq R$, we have $I_{\mathfrak{p}} = O_{\mathfrak{p}}\alpha_{\mathfrak{p}}$ with $\alpha_{\mathfrak{p}} \in B_{\mathfrak{p}}^\times$, and $\alpha_{\mathfrak{p}} = 1$ for all but finitely many $\mathfrak{p}$. Let I' be the R-lattice with $I'_{\mathfrak{p}} = O_{\mathfrak{p}}\alpha_{\mathfrak{p}}^{-1}$ for all $\mathfrak{p}$. Then I' is a left fractional O-ideal, because this is true locally. By weak approximation (Lemma 20.7.2), there exists $\beta \in B^\times$ such that $(I'\beta)_{\mathfrak{p}} = O_{\mathfrak{p}}$ for all $\mathfrak{p}$ such that $I_{\mathfrak{p}} \neq O_{\mathfrak{p}}$, i.e., for all $\mathfrak{p}$ such that $\alpha_{\mathfrak{p}} \neq 1$. But now we can perform the group operation as in Lemma 20.7.9: we have $[I]_{\mathsf{St}} + [I']_{\mathsf{St}} = [J]_{\mathsf{St}}$ where $J = I \cap I'\beta$, and for all $\mathfrak{p}$ we have

$$J_{\mathfrak{p}} = O_{\mathfrak{p}}\alpha_{\mathfrak{p}}\alpha_{\mathfrak{p}}^{-1}\beta_{\mathfrak{p}} = O_{\mathfrak{p}}\beta_{\mathfrak{p}}$$

so $J = O\beta$ and $J \simeq O$, so $[I]_{\mathsf{St}} + [I']_{\mathsf{St}} = [O]_{\mathsf{St}}$ and I' is an inverse. □

Remark 20.7.13. There is a related group to $\operatorname{StCl} O$, defined as follows. Let A be a ring, and let $\mathcal{P}(A)$ be the category of finitely generated projective left A-modules under isomorphisms. We define the group $K_0(A)$ to be the free abelian group on the isomorphism classes $[P]$ of objects $P \in \mathcal{P}(A)$ modulo the subgroup of relations

$$[P \oplus P'] = [P] + [P'], \qquad \text{for } P, P' \in \mathcal{P}(A);$$

equivalently relations $[P] + [P'] = [Q]$ for each exact sequence

$$0 \to P \to Q \to P' \to 0$$

since such a sequence splits. The group $K_0(A)$ is sometimes called the **projective class group** of A. (The group $K_0(A)$ is the *Grothendieck group* of the category $\mathcal{P}(A)$.)

Then for $P, Q \in \mathcal{P}(O)$, we have $[P] = [Q] \in K_0(O)$ if and only if P, Q are stably isomorphic [CR87, Proposition 38.22]. Moreover, there is a natural map

$$\begin{aligned} K_0(O) &\to K_0(B) \\ [P] &\mapsto [F \otimes_R O], \end{aligned}$$

and we let $SK_0(O)$ be its kernel, called the **reduced projective class group** of O. The abelian group $SK_0(O)$ is generated by elements $[P] - [Q]$ where $P, Q \in \mathcal{P}(O)$ and $F \otimes_R P \simeq F \otimes_R Q$. Finally, we have an isomorphism [CR87, Theorem 49.32]

$$\begin{aligned} \operatorname{StCl} O &\xrightarrow{\sim} SK_0(O) \\ [I]_{\mathsf{St}} &\mapsto [I] - [O]. \end{aligned} \tag{20.7.14}$$

In other words, after all of this work—at least for maximal orders—the reduced

projective class group and the stable class group coincide. (For a more general order, one instead compares to a maximal superorder via the natural extension maps $\mathrm{StCl}\, O \to \mathrm{StCl}\, O'$.)

The stable class group was first introduced and studied by Swan [Swa60, Swa62] in this context in the special case where $O = Z[G]$ is the group ring of a finite group G.

Definition 20.7.15. We say that O has **stable cancellation** (or the **simplification property**) if stable isomorphism implies isomorphism, i.e., if whenever I, I' are locally principal left O-ideals with $I \oplus O^r \simeq I' \oplus O^r$ for $r \geq 0$, then in fact $I \simeq I'$.

If we had defined stable isomorphism and cancellation for locally free O-modules, we would arrive at the same groups and condition, so stable cancellation is also called the **locally free cancellation**.

From now on, suppose that R is a global ring with $F = \mathrm{Frac}\, R$, and $O \subset B$ is a *maximal* R-order in a quaternion algebra B over F. We recall section 17.8, and the class group $Cl_\Omega R$, where $\Omega \subseteq \mathrm{Ram}\, B$ is the set of real ramified places.

Theorem 20.7.16 (Fröhlich–Swan). *Let $R = R_{(S)}$ be a global ring, let B be a quaternion algebra over F, and let $O \subset B$ be a maximal R-order. Then the reduced norm induces an isomorphism*

$$\mathrm{nrd} : \mathrm{StCl}\, O \xrightarrow{\sim} Cl_\Omega R \tag{20.7.17}$$

of finite abelian groups.

Proof. See Fröhlich [Frö75, Theorem 2, S X], Swan [Swa80, Theorem 9.4], or Curtis–Reiner [CR87, Theorem 49.32]; we will sketch a proof of a more general version of this theorem in section 28.10, when we have idelic methods at our disposal. □

20.7.18. Since B is a quaternion algebra, the notions of invertible and locally principal coincide. Then there is a surjective map of sets

$$\begin{aligned} \mathrm{Cls}_{\mathsf{L}}\, O &\to \mathrm{StCl}\, O \\ [I]_{\mathsf{L}} &\mapsto [I]_{\mathsf{St}}. \end{aligned} \tag{20.7.19}$$

Suppose further that $F = \mathrm{Frac}\, R$ is a number field and R is a global ring. Then $\mathrm{Cls}_{\mathsf{L}} O$ is a finite set, by Main Theorem 17.7.1; consequently, the stable class group $\mathrm{StCl}\, O$ is a finite abelian group. However, the map (20.7.19) of sets need not be injective.

The order O has stable cancellation if and only if the map (20.7.19) is injective (equivalently, bijective).

20.7.20. Suppose that B satisfies the Eichler condition. Then by Eichler's theorem (Theorem 17.8.3), the reduced norm also gives a bijection $\mathrm{Cls}\, O \xrightarrow{\sim} \mathrm{Cl}_\Omega R$ compatible with the surjective map $\mathrm{Cls}_{\mathsf{L}} O \to \mathrm{StCl}\, O$ (20.7.19) which must therefore also be a bijection.

What remains, then, is the case where B is definite. We restrict attention to the case where the base field F is a number field, hence a totally real field, and we work with R-orders $O \subseteq B$, where $R = \mathbb{Z}_F$ is the ring of integers of F. Vignéras [Vig76b] initiated the classification of definite quaternion orders with stable cancellation, and

showed that there are only finitely many such orders. Hallouin–Maire [HM2006] and Smertnig [Sme2015] extended this classification to certain classes of orders, and the complete classification was obtained by Smertnig–Voight [SV2019].

Theorem 20.7.21. (Vignéras, Hallouin–Maire, Smertnig, Smertnig–Voight). *Up to isomorphism, there are exactly* 316 *definite quaternion* R*-orders with stable cancellation.*

The isomorphisms in Theorem 20.7.21 are as R-orders; up to *ring* isomorphism (identifying Galois conjugates), there are exactly 247.

Example 20.7.22. If $\mathcal{O}$ is a definite maximal quaternion $\mathbb{Z}$-order, by Theorem 20.7.16 we have $\operatorname{StCl}\mathcal{O} = \#\operatorname{Cl}^+\mathbb{Z} = 1$, so $\mathcal{O}$ has stable cancellation if and only if $\#\operatorname{Cls}\mathcal{O} = 1$. These orders will be classified in section 25.4: they are the orders of discriminant $D = 2, 3, 5, 7, 13$. (More generally, if $R = \mathbb{Z}_F$ has $\#\operatorname{Cl}^+\mathbb{Z}_F = 1$, then a definite, maximal quaternion R-order has stable cancellation if and only if $\#\operatorname{Cls}\mathcal{O} = 1$.)

Remark 20.7.23. Jacobinski [Jaci68] was the first to consider the stable class group for general orders in the context of his work on genera of lattices; his cancellation theorem states more generally that if B is a central simple algebra over F and B is not a totally definite quaternion algebra, then every R-order $\mathcal{O} \subseteq B$ has stable cancellation. This result was reformulated by Fröhlich [Frö75] in terms of ideles and further developed by Fröhlich–Reiner–Ullom [FRU74]. Swan [Swa80] related cancellation to strong approximation in the context of K-groups.

Brzezinski [Brz83b] also defines the *spinor class group* of an order, a quotient of its locally free class group; this group measures certain invariants phrased in terms of quadratic forms.

Remark 20.7.24. More generally, a ring A in which every stably free right A-module is free is called a **(right) Hermite ring** by some authors: for further reference and comparison of terminology, see Lam [Lam2006, Section I.4]. If $\mathcal{O}$ has locally free cancellation, then $\mathcal{O}$ is Hermite; however, the converse does not hold in general—a counterexample is described in detail by Smertnig [Sme2015]. Smertnig–Voight [SV2019] show that there are exactly 375 definite quaternion R-orders with the Hermite property up to isomorphism.

Exercises

Throughout these exercises, let R be a noetherian domain with $F = \operatorname{Frac} R$, let B be a finite-dimensional F-algebra, let $\mathcal{O} \subseteq B$ be an R-order, and let $J = \operatorname{rad}\mathcal{O}$.

▷ 1. Let M, N be left $\mathcal{O}$-lattices.

(a) Show that a sequence $0 \to M \to N \to M' \to 0$ of left $\mathcal{O}$-lattices is exact if and only if the sequences $0 \to M_{\mathfrak{p}} \to N_{\mathfrak{p}} \to M'_{\mathfrak{p}} \to 0$ are exact for all primes $\mathfrak{p} \subseteq R$. *[Hint: Consider the modules measuring the failure of exactness and show they are locally zero, hence zero.]*

(b) Let $\phi \colon M \to N$ be a surjective O-module homomorphism. Show that ϕ splits (there exists $\psi \colon N \to M$ such that $\phi\psi = \mathrm{id}_N$) if and only if $\phi_{\mathfrak{p}} : M_{\mathfrak{p}} \to N_{\mathfrak{p}}$ splits for all primes $\mathfrak{p} \subseteq R$.

▹ 2. Suppose R is a DVR and B is a quaternion algebra. Let $J = \operatorname{rad} O$. Show that $\overline{J} = J$ and $O_{\mathsf{L}}(\operatorname{rad} O) = O_{\mathsf{R}}(\operatorname{rad} O)$.

3. Let R be a complete DVR with $\mathfrak{p} = \operatorname{rad} R$. Show that the $\mathfrak{p}$-adic topology and the J-adic topology on O are the same.

4. Let R be a DVR with maximal ideal $\mathfrak{p}$, and let $O = \begin{pmatrix} R & R \\ \mathfrak{p} & R \end{pmatrix} \subseteq B = \mathrm{M}_2(F)$. Let $I \subseteq B$ be a left fractional O-ideal. Show that either I is invertible as a O-ideal or I is conjugate to $\mathrm{M}_2(R)$ by an element of $B^\times$.

5. Let O be a maximal R-order, and let M be a projective left O-lattice. Show that M is indecomposable if and only if FM is a simple left B-module. *[Hint: Suppose $W \subseteq FM$ is a left B-submodule of FM, and let $N := M \cap W$. Show that M/N is a projective O-lattice, so the sequence $0 \to N \to M \to M/N \to 0$ splits.]*

▹ 6. Let A be a ring (not necessarily commutative, but with 1. In this exercise, we prove Lemma 20.4.8, that

$$\operatorname{rad} A = \{\beta \in A : 1 - \alpha_1\beta\alpha_2 \in A^\times \text{ for all } \alpha_1, \alpha_2 \in A\}.$$

We first show the inclusion ($\subseteq$).

(a) Since $\operatorname{rad} A$ is a two-sided ideal, it suffices to show that $1 - \beta \in A^\times$. Show that $A(1-\beta) = A$.

(b) Let $\alpha \in A$ be such that $\alpha(1-\beta) = 1$. Repeating the argument, show that $A(1-(1-\alpha)) = A\alpha = A$.

(c) Show that α is also a right inverse of $1-\beta$, so $1-\beta \in A^\times$.

Next we show the inclusion ($\supseteq$).

(d) Let $\beta \in A$ be such that $1 - \alpha\beta\gamma \in A^\times$ for all $\alpha, \gamma \in A$. Let M be a simple left A-module. Show that $\alpha M = \{0\}$. Conclude that $\alpha \in \operatorname{rad} A$.

▹ 7. Suppose R is a complete DVR. Prove Lemma 20.6.8: the association $I \mapsto I/JI$ gives a bijection between isomorphism classes of indecomposable finitely generated projective left O-modules and isomorphism classes of simple finite-dimensional left O/J-modules.

▹ 8. Let R be a complete DVR, and let I, I', J be finitely generated left O-modules such that

$$I \oplus J \simeq I' \oplus J$$

as left O-modules. Prove that $I \simeq I'$ as left O-modules.

9. Let $\Lambda = \mathrm{M}_n(O)$ with $n \geq 1$. Show that $\operatorname{StCl}\Lambda \simeq \operatorname{StCl} O$.
10. Show that the Lipschitz order has stable cancellation.
11. Let $B = (-1, -3 \mid \mathbb{Q})$ and let $O = \mathbb{Z} + \mathbb{Z}(3i) + \mathbb{Z}(-1 + j)/2 + \mathbb{Z}(3i + ij)/2$.
    (a) Show that O is an order with $\operatorname{discrd} O = 9$.
    (b) Show that O has stable cancellation.

This chapter is licensed under the terms of the Creative Commons Attribution-NonCommercial 4.0 International License (http://creativecommons.org/licenses/by-nc/4.0/), which permits any noncommercial use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license and indicate if changes were made.

The images or other third party material in this chapter are included in the chapter's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the chapter's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.

Chapter 21

Hereditary and extremal orders

In this chapter, we consider hereditary orders, those with the simplest kind of module theory; we characterize these orders in several ways, including showing they have an extremal property with respect to their Jacobson radical.

21.1 ▹ Hereditary and extremal orders

Let R be a Dedekind domain. Then R is **hereditary**: every submodule of a projective module is again projective. (Hence the name: projectivity is inherited by a submodule.) A noetherian domain is hereditary if and only if every ideal of R is projective, or equivalently, that every submodule of a free R-module is a direct sum of ideals of R. This property is used in the proof of unique factorization of ideals and makes the structure theory of modules over a Dedekind quite nice. (Note, however, that every order in a number field which is not maximal is not hereditary.)

It is important to identify those orders for which projective modules abound. Let B be a simple finite-dimensional F-algebra and let $\mathcal{O} \subseteq B$ be an R-order.

Definition 21.1.1. We say $\mathcal{O}$ is **left hereditary** if every left $\mathcal{O}$-ideal $I \subseteq \mathcal{O}$ is projective as a left $\mathcal{O}$-module.

We could define also *right hereditary*, but left hereditary and right hereditary are equivalent for an R-order $\mathcal{O}$, and so we simply say **hereditary**. We have $\mathcal{O}$ hereditary if and only if every $\mathcal{O}$-submodule of a projective finitely generated $\mathcal{O}$-module is projective—that is to say, projectivity is *inherited* by submodules. Moreover, being hereditary is a local property.

Maximal orders are hereditary (Theorem 18.1.2), and one motivation for hereditary orders is that many of the results from chapter 18 on the structure of two-sided ideals extend from maximal orders to hereditary orders (Theorem 21.4.9).

Proposition 21.1.2. *Suppose $\mathcal{O}$ is hereditary. Then the set of two-sided invertible fractional $\mathcal{O}$-ideals of B forms an abelian group under multiplication, generated by the prime $\mathcal{O}$-ideals.*

© The Author(s) 2021

J. Voight, *Quaternion Algebras*, Graduate Texts in Mathematics 288,
https://doi.org/10.1007/978-3-030-56694-4_21

Hereditary orders are an incredibly rich class of objects, and they may be characterized in a number of equivalent ways (Theorem 21.5.1). We restrict to the complete local case, and suppose now that R is a complete DVR with unique maximal ideal $\mathfrak{p}$ and residue field $k = R/\mathfrak{p}$.

Just as maximal orders are defined in terms of containment, we say O is **extremal** if whenever $O' \supseteq O$ and $\operatorname{rad} O' \supseteq \operatorname{rad} O$, then $O' = O$. If O is not extremal, then

$$O' := O_{\mathsf{L}}(\operatorname{rad} O) \supsetneq O \tag{21.1.3}$$

is a superorder. We then have the following main theorem (Theorem 21.5.1).

Main Theorem 21.1.4. *Let R be a complete DVR and let $O \subseteq B$ be an R-order in a simple F-algebra B. Let $J := \operatorname{rad} O$. Then the following are equivalent:*

(i) *O is hereditary;*
(ii) *J is projective as a left O-module;*
(ii′) *J is projective as a right O-module;*
(iii) *$O_{\mathsf{L}}(J) = O$;*
(iii′) *$O_{\mathsf{R}}(J) = O$;*
(iv) *J is invertible as a (sated) two-sided O-ideal; and*
(v) *O is extremal.*

The fact that hereditary orders are the same as extremal orders is quite remarkable, and gives tight control over the structure of hereditary orders: extremal orders are equivalently characterized as endomorphism algebras of flags in a suitable sense, and so we have the following important corollary for quaternion algebras.

Corollary 21.1.5. *Suppose further that B is a quaternion algebra. Then an R-order $O \subseteq B$ is hereditary if and only if either O is maximal or*

$$O \simeq \begin{pmatrix} R & R \\ \mathfrak{p} & R \end{pmatrix} \subseteq \mathrm{M}_2(F) \simeq B.$$

It is no surprise that we meet again the order from Example 20.1.2! The reader who is willing to accept Corollary 21.1.5 can profitably move on from this chapter, as the ring of upper triangular matrices is explicit enough to work with in many cases. That being said, the methods we encounter here will be useful in framing investigations of orders beyond the hereditary ones.

21.2 Extremal orders

In this section, we will see how to extend an order to a superorder using the Jacobson radical, and we will characterize those orders that are extremal with respect to this process.

We work locally throughout this section; let R be a complete DVR with maximal ideal $\mathfrak{p} = \operatorname{rad}(R)$ and residue field $k = R/\mathfrak{p}$, and let $F = \operatorname{Frac} R$. Let B be a finite-dimensional separable F-algebra and let $O \subseteq B$ be an R-order.

21.2.1. Our motivation comes from the following: we canonically associate a superorder as follows. Let $J := \operatorname{rad} O$ and $O' := O_{\mathsf{L}}(J)$. Then $O' \supseteq O$. By Corollary 20.5.5, $J^r \subseteq \mathfrak{p}O \subseteq \mathfrak{p}O'$ for some $r > 0$, and then $J \subseteq \operatorname{rad} O'$.

Definition 21.2.2. An R-order $O' \subseteq B$ **radically covers** O if $O' \supseteq O$ and $\operatorname{rad} O' \supseteq \operatorname{rad} O$. We say O is **extremal** if whenever O' radically covers O then $O' = O$.

We can think of extremal orders as like maximal orders, but under certain inclusions.

Proposition 21.2.3. *An R-order O is extremal if and only if $O_{\mathsf{L}}(\operatorname{rad} O) = O$ if and only if $O_{\mathsf{R}}(\operatorname{rad} O) = O$.*

Proof. The argument is due to Jacobinski [Jaci71, Proposition 1].

We first prove ($\Rightarrow$). Suppose O is extremal, and let $J = \operatorname{rad} O$ and $O' = O_{\mathsf{L}}(J)$. By Corollary 20.5.5, J is topologically nilpotent as a O-ideal, so the same is true as a O' ideal, and $J \subseteq \operatorname{rad} O'$ and O' radically covers O. Since O is extremal, we conclude $O' = O$. The same argument works on the right.

Next we prove ($\Leftarrow$). Let $J = \operatorname{rad} O$, suppose $O = O_{\mathsf{L}}(J)$; let O' radically cover O, and let $J' = \operatorname{rad} O'$. As lattices, we have $\mathfrak{p}^s O' \subseteq J$ for some $s > 0$; by Theorem 20.5.1, $(J')^r \subseteq \mathfrak{p}O'$ for some $r > 0$, so putting these together we have $(J')^t \subseteq J$ for some $t > 0$. Suppose $t > 1$. Since O' radically covers, we have $J \subseteq J'$; thus $J(J')^{t-1} \subseteq (J')^t \subseteq J$ and $(J')^{t-1} \subseteq O_{\mathsf{R}}(J) = O$. But then since $((J')^{t-1})^t \subseteq (J')^t \subseteq J$, by Corollary 20.5.5, $(J')^{t-1} \subseteq J$. Continuing in this way, we obtain $t = 1$ and $J' \subseteq J$. Therefore $J = J'$ and $O = O_{\mathsf{L}}(J) = O_{\mathsf{L}}(J') = O'$, thus O is extremal. □

Lemma 21.2.4. *Let O be an R-order and let $O' \subseteq B$ be an R-order containing O. Let $J' := \operatorname{rad} O'$. Then $O + J'$ is an R-order that radically covers O. If further $J' \subseteq O$, then $J' \subseteq J$.*

Proof. See Exercise 21.6. □

21.2.5. In view of Lemma 21.2.4, an extremal order is determined by its homomorphic image in a nice k-algebra as follows.

Let O be an extremal R-order and let $O' \subseteq B$ be a maximal R-order containing O. Let $J' := \operatorname{rad} O'$. By Lemma 21.2.4, $O + J'$ is an R-order that radically covers O, so $O + J' = O$. Therefore $J' \subseteq O$. By the second part of Lemma 21.2.4, we immediately conclude $J' \subseteq J := \operatorname{rad} O$. In sum,

$$J' = \operatorname{rad} O' \subseteq J = \operatorname{rad} O \subseteq O. \tag{21.2.6}$$

Consider now the reduction map $\rho\colon O' \to O'/J'$. Since $J' \subseteq O$, if $A = \rho(O)$ then $O = \rho^{-1}(A)$. But since $\mathfrak{p}O' \subseteq J'$ and O' is a maximal R-order, the codomain is a nice, finite dimensional k-algebra, something we will get our hands on in the next section.

Paragraph 21.2.5 has the following consequence.

Lemma 21.2.7. *Suppose that B is a division algebra over F and let $O \subseteq B$ be extremal. Then O is maximal.*

Proof. Recall 13.3.7. The valuation ring $O' \supseteq O$ has the unique maximal two-sided ideal $J' = \operatorname{rad} O' = O' \setminus (O')^\times$, so O'/J' is a field. We have (21.2.6) $J' \subseteq J$, but then $J/J' = \{0\} \subseteq O'/J'$ thus $J = J'$. Thus O' radically covers O, and since O is extremal, $O = O'$. □

Remark 21.2.8. We stop short in our explicit description of local extremal orders in section 21.2: we gave a construction in 21.3.1 only for $B \simeq \mathrm{M}_n(F)$. The results extend to $B \simeq \mathrm{M}_n(D)$ where D is a division algebra over F by considering lattices in a free left D-module: see Reiner [Rei2003, Theorem 39.14].

21.3 ∗ Explicit description of extremal orders

We now turn to an explicit description of extremal orders. In Lemma 10.5.4, we saw that maximal orders in a matrix algebra $B = \operatorname{End}_F(V)$ are endomorphism algebras of lattices. In this section, we extend this to encompass orders that arise from endomorphism algebras of a chain of lattices: these orders are "block upper triangular", and can be characterized in a number of ways.

21.3.1. Let V be a finite-dimensional F-vector space and let $B = \operatorname{End}_F(V)$; then V is a simple B-module. Let $M \subseteq V$ be an R-lattice. By Lemma 10.5.4, $\Lambda := \operatorname{End}_R(M)$ is a maximal R-order; we have $\operatorname{rad}\Lambda = \mathfrak{p}\Lambda$.

Choosing a basis for M, we get $\Lambda \simeq \mathrm{M}_n(R) \subseteq \mathrm{M}_n(F) \simeq B$, and $\operatorname{rad}\Lambda = \mathrm{M}_n(\mathfrak{p})$.

Now let $Z := M \otimes_R k = M/\mathfrak{p}M$. Then Z is a finite-dimensional vector space over k. Let

$$\mathcal{E}\colon \{0\} = Z_0 \subsetneq Z_1 \subsetneq \cdots \subsetneq Z_{s-1} \subsetneq Z_s = Z$$

be a **(partial) flag**, a strictly increasing sequence of k-vector spaces. We define

$$O_{\mathsf{L}}(\mathcal{E}) := \{\alpha \in \Lambda : \alpha Z_i \subseteq Z_i : i = 0, \dots, s\}.$$

Equivalently, let M_i be the inverse image of Z_i under the projection $M \to Z$; then we have a chain

$$\mathfrak{p}M = M_0 \subsetneq M_1 \subsetneq \cdots \subsetneq M_{s-1} \subsetneq M_s = M \tag{21.3.2}$$

and

$$O_{\mathsf{L}}(\mathcal{E}) = \{\alpha \in \Lambda : \alpha M_i \subseteq M_i : i = 0, \dots, s\}.$$

Lemma 21.3.3. *$O_{\mathsf{L}}(\mathcal{E}) \subseteq \Lambda$ is an R-order with*

$$\operatorname{rad} O_{\mathsf{L}}(\mathcal{E}) = \{\alpha \in \Lambda : \alpha Z_i \subseteq Z_{i-1}\} = \{\alpha \in \Lambda : \alpha M_i \subseteq M_{i-1}\}.$$

Proof. That $O_{\mathsf{L}}(\mathcal{E})$ is an order follows in the same way as the proof of Lemma 10.2.7. For the statement on the radical: let $J = \{\alpha \in \Lambda : \alpha Z_i \subseteq Z_{i-1}\}$. Then $J \subseteq O_{\mathsf{L}}(\mathcal{E})$ is a two-sided ideal. We have $J^s \subseteq \mathfrak{p}\Lambda$ pushing along the flag, so $J \subseteq \operatorname{rad} O_{\mathsf{L}}(\mathcal{E})$ by Corollary 20.5.5. Conversely,

$$O_{\mathsf{L}}(\mathcal{E})/J \simeq \bigoplus_{i=1}^{s} \operatorname{End}_k(Z_i/Z_{i-1});$$

each factor is simple, so the sum is (Jacobson) semisimple; therefore $J \supseteq \operatorname{rad} O_{\mathsf{L}}(\mathcal{E})$ and equality holds. □

Example 21.3.4. If we take the trivial flag $\mathcal{O}_\mathsf{L}(\mathcal{E}) : \{0\} = Z_0 \subsetneq Z_1 = Z$, then $\mathcal{O}_\mathsf{L}(\mathcal{E}) = \Lambda$, so this recovers the construction of maximal orders.

Example 21.3.5. Let $\mathcal{E}$ be the **complete** flag of length $s = n+1 = \dim_F V$, where each quotient has $\dim_k(Z_{i+1}/Z_i) = 1$. Then there exists a basis $z_1, \dots, z_n$ of Z so that Z_i has basis $z_1, \dots, z_{n-i}$; We lift this to basis to $x_1, \dots, x_n$ of M (by Nakayama's lemma), and in this basis, we have

$$\mathcal{O}_\mathsf{L}(\mathcal{E}) = \begin{pmatrix} R & R & R & \dots & R \\ \mathfrak{p} & R & R & \dots & R \\ \mathfrak{p} & \mathfrak{p} & R & \dots & R \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \mathfrak{p} & \mathfrak{p} & \mathfrak{p} & \dots & R \end{pmatrix}$$

consisting of matrices which are upper triangular modulo $\mathfrak{p}$, and

$$\operatorname{rad} \mathcal{O}_\mathsf{L}(\mathcal{E}) = \begin{pmatrix} \mathfrak{p} & R & R & \dots & R \\ \mathfrak{p} & \mathfrak{p} & R & \dots & R \\ \mathfrak{p} & \mathfrak{p} & \mathfrak{p} & \dots & R \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \mathfrak{p} & \mathfrak{p} & \mathfrak{p} & \dots & \mathfrak{p} \end{pmatrix} = \mathcal{O}_\mathsf{L}(\mathcal{E}) \begin{pmatrix} 0 & 1 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & 1 \\ \pi & 0 & \dots & 0 \end{pmatrix} \tag{21.3.6}$$

where the latter is taken to be a block matrix with lower left entry π and top right entry the $(n-1) \times (n-1)$ identity matrix.

Other choices of flag give an order which lie between $\mathcal{O}_\mathsf{L}(\mathcal{E})$ and Λ: we might think of them as being *block upper triangular* orders.

Now for the punch line of this section.

Proposition 21.3.7. *Let $\mathcal{O} \subseteq B$ be an R-order. Then $\mathcal{O}$ is extremal if and only if $\mathcal{O} = \mathcal{O}_\mathsf{L}(\mathcal{E})$ for a flag $\mathcal{E}$.*

Proof. Let $\mathcal{O} = \mathcal{O}_\mathsf{L}(\mathcal{E})$. Let $J = \operatorname{rad} \mathcal{O}$; we seek to apply Proposition 21.2.3, so we show that $\mathcal{O} = \mathcal{O}_\mathsf{L}(J)$. By Lemma 21.3.3, we have $JM_i = M_{i-1}$ so $\mathcal{O}_\mathsf{L}(J)M_{i-1} = \mathcal{O}_\mathsf{L}(J)JM_i = JM_i = M_{i-1}$ for $i = 1, \dots, s$. Since $M_0 = \mathfrak{p}M \simeq M$, we conclude $\mathcal{O}_\mathsf{L}(J) = \mathcal{O}_\mathsf{L}(\mathcal{E}) = \mathcal{O}$ by definition.

Conversely, suppose $\mathcal{O}$ is extremal with $J = \operatorname{rad} \mathcal{O}$. Let s be minimal so that $J^s = \mathfrak{p}\mathcal{O}$. We may embed $\mathcal{O} \subseteq \Lambda$ for some Λ, and we take the flag

$$\mathcal{E}\colon \{0\} = J^s Z \subsetneq J^{s-1} Z \subsetneq \cdots \subsetneq JZ \subsetneq Z.$$

Then $\mathcal{O} \subseteq \mathcal{O}(\mathcal{E})$ and $\operatorname{rad} \mathcal{O}(\mathcal{E}) \supseteq J$ by construction, so since $\mathcal{O}$ is extremal, we have $\mathcal{O} = \mathcal{O}(\mathcal{E})$. ∎

21.4 Hereditary orders

We now link the orders in the previous two sections to another important type of order. The theory of extremal and hereditary orders was developed by Brumer [Brum63a, Brum63b], Drozd–Kirichenko [DK68], Harada [Har63a, Har63b, Har63c], Jacobinski [Jaci71], and Hijikata–Nishida [HN94]. An overview of the local and global theory of hereditary orders is given by Reiner [Rei2003, §§2f, 39–40], and Drozd–Kirichenko–Roiter [DKR67] and Hijikata–Nishida [HN94] extend some results from hereditary orders to Bass orders.

Let R be a noetherian domain with $F = \operatorname{Frac} R$, and let B be a separable F-algebra, and let $O \subseteq B$ be an R-order.

Definition 21.4.1. We say O is **left hereditary** if every left O-ideal $I \subseteq O$ is projective as a left O-module.

21.4.2. We could similarly define *right hereditary*, but since an order O is left and right noetherian, it follows that O is left hereditary if and only if O is right hereditary: see Exercise 21.8. When B is a quaternion algebra, the standard involution interchanges and left and right, so the two notions are immediately seen to be equivalent. Accordingly, we say **hereditary** for either sided notion.

Example 21.4.3. In the generic case $F = R$ and $O = B$, we note that every semisimple algebra B over a field F is hereditary: by Lemma 7.3.5, every B-module is semisimple hence the direct sum of simple B-modules equivalently maximal left ideals, by Lemma 7.2.7.

21.4.4. By 20.2.6, being hereditary is a local property.

The following lemma motivates the name *hereditary*: projectivity is inherited by submodules. (Note that since R is noetherian, a finitely generated O-module is noetherian, so every submodule is finitely generated.)

Lemma 21.4.5. *Let O be hereditary, and let P be a finitely generated projective left O-module. Then every submodule $M \subseteq P$ is isomorphic as a left O-module to a finite direct sum of finitely generated left O-ideals; in particular, M is projective.*

Proof. We may suppose without loss of generality that $P \simeq O^r$. We proceed by induction on r; the case $r = 1$ holds by definition. Decompose $O^r = E \oplus O$ where $E \simeq O^{r-1}$. From the exact sequence

$$0 \to \ker\phi \to M \to M \cap E \to 0$$

and projectivity, we find that $M \simeq (M \cap E) \oplus \ker\phi$ where $\ker\phi \subseteq O$ is a left ideal of O. By induction, $M \cap E$ is projective, so the same is true of M. □

Corollary 21.4.6. *O is hereditary if and only if every submodule of a projective O-module is projective.*

Proof. The implication (⇒) is Lemma 21.4.5; for the implication (⇐), O is projective (free!) as a left O-module and every left ideal is a O-submodule $I \subseteq O$, so by hypothesis I is projective. □

Remark 21.4.7. O is hereditary if and only if every R-lattice $I \subseteq B$ with $O \subseteq O_{\mathsf{L}}(I)$ is projective as a left O-module, after rescaling.

However, a bit of a warning is due. If $I \subseteq B$ is an R-lattice that is projective as a left O-module, then we have shown that I is projective as a left $O_{\mathsf{L}}(I)$-module (Lemma 20.3.1), whence left invertible (Theorem 20.3.3) as a lattice. But the converse need not be true; so it is important that in the definition of hereditary we do not require that every left O-fractional ideal I is invertible *as a left fractional O-ideal* (Definition 16.5.17): the latter carries the extra assumption that I is sated. See also Remark 20.3.8.

Lemma 21.4.8. *Let $O \subseteq B$ be a hereditary R-order and let $O' \supseteq O$ be an R-superorder. Then O' is hereditary.*

Proof. Let $I' \subseteq O'$ be a left O'-ideal. Scaling we may take $I' \subseteq O$, and it is a left O-ideal. Since O is hereditary, I' is projective as a left O-module; by Lemma 20.3.1, I' is projective as a left O'-module. □

One of the desirable aspects of hereditary orders is that many of the results from chapter 18 on the structure of two-sided ideals extend from maximal orders to hereditary orders. Indeed, section 18.2 made no maximality hypothesis (we held out as long as we could!).

Theorem 21.4.9. *Let R be a Dedekind domain and let O be a hereditary R-order in a simple F-algebra B. Then the set of two-sided invertible fractional O-ideals of B forms an abelian group under multiplication, generated by the invertible prime O-ideals.*

Proof. Proven in the same manner as in Theorem 18.3.4; a self-contained proof is requested in Exercise 21.4. □

Remark 21.4.10. Theorem 21.4.9 is proven by Vignéras [Vig80a, Théorème I.4.5], but there is a glitch in the proof. Let R be a Dedekind domain, let B be a quaternion algebra over $F = \operatorname{Frac} R$, and let $O \subseteq B$ be an R-order. Vignéras claims that the two-sided ideals of O form a group that is freely generated by the *prime* ideals, and the proof uses that if I is a two-sided ideal then I is invertible. This is false for a general order O (see Example 16.5.12).

If one restricts to the group of invertible two-sided ideals, the logic of the proof is still flawed. The proof does not use anything about quaternion algebras, and works verbatim for the case where $R = \mathbb{Z} \subseteq F = \mathbb{Q}$ and B is replaced by $K = \mathbb{Q}(\sqrt{d_K})$ and O is replaced by an order of discriminant $d = d_K f^2$ that is not maximal, of conductor $f \in \mathbb{Z}_{>1}$, as in section 16.1. Then the ideal $\mathfrak{f} = f\mathbb{Z} + \sqrt{d}\mathbb{Z}$ is not invertible, but $\mathfrak{f} \supsetneq (f)$ and (f) is invertible but not maximal, so the group of invertible ideals is not generated by primes.

However, if one supposes that *every* two-sided ideal is invertible (as a lattice), then the argument can proceed: this is the class of hereditary orders, and is treated in Theorem 21.4.9.

Remark 21.4.11. The module theory for hereditary noetherian prime rings, generalizing hereditary orders, has been worked out by Levy–Robson [LR2011].

21.5 ∗ Classification of local hereditary orders

We now come to the main theorem of this chapter, relating extremal orders, hereditary orders, their modules and composition series in the local setting.

Theorem 21.5.1. *Let R be a complete DVR and let $F := \operatorname{Frac} R$. Let B be a finite-dimensional F-algebra, and let $\mathcal{O} \subseteq B$ be an R-order. Let $J := \operatorname{rad} \mathcal{O}$. Then the following are equivalent, along with the conditions′ where 'left' is replaced by 'right':*

(i) *$\mathcal{O}$ is extremal;*
(ii) *Every projective indecomposable left $\mathcal{O}$-submodule $P \subseteq B$ is the minimum $\mathcal{O}$-supermodule of JP;*
(iii) *Every projective indecomposable left $\mathcal{O}$-module P has a unique composition series;*
(iv) *Every projective indecomposable left $\mathcal{O}$-module P has a unique composition series consisting of projectives;*
(v) *$\mathcal{O}$ is hereditary;*
(vi) *J is projective as a left $\mathcal{O}$-module;*
(vii) *If P is a projective indecomposable left $\mathcal{O}$-module, then JP is also projective indecomposable; and*
(viii) *J is invertible as a (sated) two-sided $\mathcal{O}$-ideal.*

Proof. See Hijikata–Nishida [HN94, §1]. □

Corollary 21.5.2. *A maximal order is hereditary.*

Proof. We proved this in Theorem 18.1.2, but here is another proof using Theorem 21.5.1: the property of being maximal is local, and a maximal order is extremal. □

To conclude, we classify the lattices of a local hereditary order.

21.5.3. Suppose R is a complete DVR and that $B \simeq \mathrm{M}_n(F)$. Suppose $\mathcal{O} \subseteq B$ is a hereditary R-order; then by Theorem 21.5.1, $\mathcal{O} = \mathcal{O}_{\mathsf{L}}(\mathcal{E})$ is extremal, arising from a chain 21.3.2 which by Lemma 21.3.3 is of the form

$$\mathfrak{p}M = M_0 = J^s M \subsetneq J^{s-1} M \subsetneq \cdots \subsetneq M_{s-1} = JM \subsetneq M_s = M,$$

with each quotient $M_i/M_{i+1} \simeq M/JM$ simple.

We claim that the set $M, JM, \dots, J^{s-1}M$ form a complete set of isomorphism classes of indecomposable left $\mathcal{O}$-modules. Indeed, these modules are all mutually nonisomorphic, because an isomorphism $\phi\colon J^i M \xrightarrow{\sim} J^j M$ of left $\mathcal{O}$-modules extends to an isomorphism $\phi \in \operatorname{End}_B(B) \simeq F$ so is given by (right) multiplication by a power of π, impossible unless $i \equiv j \pmod{s}$. And if N is an indecomposable left $\mathcal{O}$-module, then $FN \simeq F^n$ is 'the' simple B-module, so N is isomorphic to a lattice in V. Since J is invertible, we may replace N by $J^r N$ with $r \in \mathbb{Z}$, to suppose that $M \supseteq N \supsetneq JM$. But $M/JM \simeq \mathcal{O}/J$ is simple as a left $\mathcal{O}$-module, so $N = M$. (See also Reiner [Rei2003, Theorem 39.23].)

Exercises

1. Show that a Dedekind domain is hereditary (cf. Exercise 9.5).
2. Let $R = \mathbb{Z}$, let $B = K = \mathbb{Q}(\sqrt{d})$ with d the discriminant of K, and let $S \subseteq \mathbb{Z}_K$ be an order. Show that S is hereditary if and only if S is maximal.
3. Let R be a DVR with maximal ideal $\mathfrak{p} = \pi R$ and $F = \operatorname{Frac} R$ with $\operatorname{char} F \neq 2$. Let $B = \left(\dfrac{1, \pi}{F}\right)$ and $O = R\langle i, j\rangle$ the standard order. Show directly that $\operatorname{rad} O = Oj = jO$, and conclude that O is hereditary (but not a maximal order).
4. ▷ Give a self-contained proof of Theorem 21.4.9 following Theorem 18.3.4. (Where does the issue with invertibility arise?)
5. Let R be a complete DVR and let O be a hereditary R-order. Show that O is hereditary if and only if $\operatorname{rad} O$ is an invertible (sated) two-sided O-ideal.
6. ▷ In this exercise, we prove Lemma 21.2.4 following Reiner [Rei2003, Exercise 39.2]. We adopt the notation from that section, so in particular R be a complete DVR with maximal ideal $\mathfrak{p} = \operatorname{rad}(R)$. Let O be an R-order and let $O' \subseteq B$ be an R-order containing O. Let $J' = \operatorname{rad} O'$.

   (a) Show that $O + J'$ is an R-order.
   (b) Show that $O + J'$ radically covers O. *[Hint: let $J = \operatorname{rad} O$, and claim that $J + J' \subseteq \operatorname{rad}(O + J')$. For r large, show $J^r \subseteq \mathfrak{p}O$ so $(J + J')^r \subseteq \mathfrak{p}O' + J'$ and $(J')^r \subseteq \mathfrak{p}O'$, and then making r even larger show $(J + J')^{r^3} \subseteq \mathfrak{p}(O + J')$. Conclude using Corollary 20.5.5.]*
   (c) If further $J' \subseteq O$, show that $J' \subseteq J$.

7. Let R be a Dedekind domain with $F = \operatorname{Frac}(R)$, let B be finite-dimensional F-algebra, and let $O \subseteq B$ be a hereditary order. Let P be a finitely generated projective O-module. Show that P is indecomposable if and only if $V := P \otimes_R F$ is simple as a B-module.
8. ▷ Let R be a Dedekind domain, and let $O \subseteq B$ be an R-order in a finite-dimensional F-algebra. Show that O is left hereditary (every left O-ideal is projective) if and only if it is right hereditary (every right O-ideal is projective). [See Reiner [Rei2003, Theorem 40.1].]
9. Consider the ring
$$A := \left\{ \begin{pmatrix} a & 0 \\ b & c \end{pmatrix} : a \in \mathbb{Z}, b, c \in \mathbb{Q} \right\}.$$
Show that every submodule of a projective left A-module is projective, but the same is not true on the right.
10. Let R be a Dedekind domain. Let B be a separable F-algebra, and let $B \simeq B_1 \times \cdots \times B_r$ be its decomposition into simple components, with $B_i = Be_i$ for central idempotents e_i. Let K_i be the center of B_i, and let S_i be the integral closure of R in K_i.

    (a) Let $O \subseteq B$ be a hereditary R-order. Show that $O \simeq O_1 \times \cdots \times O_r$ where $O_i = Oe_i$, and each O_i is a hereditary R-order in B_i.

(b) Conversely, if $\mathcal{O}_i \subseteq B_i$ is a hereditary R-order, then $\mathcal{O}_1 \times \cdots \times \mathcal{O}_r$ is a hereditary R-order in B.

[Hint: use the fact that hereditary orders are extremal.]

11. For the following exercise, we consider integral group rings. Let G be a finite group of order $n = \#G$ and let R be a Dedekind domain with $F = \operatorname{Frac} R$. Suppose that $\operatorname{char} F \nmid n$. Then $B := F[G]$ is a separable F-algebra by Exercise 7.16. Let $\mathcal{O} = R[G]$.

(a) Let $\mathcal{O}' \supseteq \mathcal{O}$ be an R-superorder of $\mathcal{O}$ in B. Show that

$$\mathcal{O} \subseteq \mathcal{O}' \subseteq n^{-1}\mathcal{O}.$$

[Hint: for ell $\alpha = \sum_g a_g g \in \mathcal{O}'$ with $a_g \in F$, show that

$$\operatorname{Tr}_{B|F}(\alpha g) = n a_g \in R.$$

Conclude that $\mathcal{O}' \subseteq n^{-1}\mathcal{O}$.]

(b) Show that $\mathcal{O}$ is maximal if and only if $\mathcal{O}$ is hereditary if and only if $n \in R^\times$. *[Hint: if $\mathcal{O}$ is hereditary, then $\mathcal{O}$ contains the central idempotent $n^{-1}\sum_{g\in G} g$ by Exercise 21.10.]*

(c) We define the **left conductor** of $\mathcal{O}'$ into $\mathcal{O}$ to be the colon ideal

$$(\mathcal{O}' : \mathcal{O})_{\mathsf{L}} = \{\alpha \in B : \alpha\mathcal{O}' \subseteq \mathcal{O}\}.$$

(and similarly on right). Prove that

$$(\mathcal{O}' : \mathcal{O})_{\mathsf{L}} = \sum_{i=1}^{t} \frac{n}{n_i} \operatorname{codiff}(\mathcal{O}_i').$$

12. Give an explicit description like Example 21.3.5 for $\mathcal{O}_{\mathsf{L}}(\mathcal{E})$ when $\dim_F V = 3, 4$.

13. Let R be a Dedekind domain, and let $\mathcal{O} \subseteq B$ be an R-order in a finite-dimensional simple F-algebra. Show that $\mathcal{O}$ is maximal if and only if $\mathcal{O}$ is hereditary and $\operatorname{rad}\mathcal{O} \subseteq \mathcal{O}$ is a maximal two-sided ideal.

This chapter is licensed under the terms of the Creative Commons Attribution-NonCommercial 4.0 International License (http://creativecommons.org/licenses/by-nc/4.0/), which permits any noncommercial use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license and indicate if changes were made.

The images or other third party material in this chapter are included in the chapter's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the chapter's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.

Chapter 22

Quaternion orders and ternary quadratic forms

In this chapter, we classify orders over a Dedekind domain in terms of ternary quadratic forms; this is the integral analogue to what we did over fields in Chapter 5.

22.1 ▹ Quaternion orders and ternary quadratic forms

We begin our project by returning to the classification over fields: in Chapters 5 and 6 (see Main Theorem 5.2.5 and Theorem 6.4.7), we saw that quaternion algebras over a field F are classified by similarity classes of nondegenerate ternary quadratic forms over F. We will soon see that, suitably interpreted, quaternion orders are classified by similarity classes of integral ternary quadratic forms.

Let R be a PID with field of fractions $F := \operatorname{Frac} R$. We recall that the similarity class of a ternary quadratic form $Q\colon R^3 \to R$ is determined by the natural change of variable by $\mathrm{GL}_3(R)$ on the domain and by rescaling by $R^\times$ on the codomain, and that Q is *nondegenerate* if and only if $\operatorname{disc}(Q) \neq 0$.

Main Theorem 22.1.1. *Let R be a PID. Then there is a (reduced) discriminant-preserving bijection*

$$\left\{\begin{array}{c}\text{Nondegenerate ternary quadratic} \\ \text{forms } Q \text{ over } R \text{ up to similarity}\end{array}\right\} \leftrightarrow \left\{\begin{array}{c}\text{Quaternion orders over } R \\ \text{up to isomorphism}\end{array}\right\}.$$

One beautiful feature of the bijection in Main Theorem 22.1.1 is that it can be given explicitly. Let $Q\colon R^3 \to R$ be a ternary quadratic form with nonzero discriminant, and let e_1, e_2, e_3 be the standard basis for R^3. Then the extension to F given by $Q_F : F^3 \to F$ is a ternary quadratic space whose even Clifford algebra (section 5.3) is a quaternion algebra B. Moreover, the R-lattice $\mathcal{O}$ with basis

$$1, \quad i := e_2 e_3, \quad j := e_3 e_1, \quad k := e_1 e_2$$

is closed under multiplication and so defines an R-order in B. Explicitly, if the quadratic form Q is given by

$$Q(x, y, z) = ax^2 + by^2 + cz^2 + uyz + vxz + wxy \in R[x, y, z]$$

© The Author(s) 2021

J. Voight, *Quaternion Algebras*, Graduate Texts in Mathematics 288,
https://doi.org/10.1007/978-3-030-56694-4_22

with (half-)discriminant

$$N := 4abc + uvw - au^2 - bv^2 - cw^2 \neq 0,$$

then we associate the quaternion R-order $\mathcal{O} \subseteq B$ with basis $1, i, j, k$ and multiplication laws

$$\begin{aligned} i^2 &= ui - bc & jk &= a\bar{i} = a(u-i) \\ j^2 &= vj - ac & ki &= b\bar{j} = b(v-j) \\ k^2 &= wk - ab & ij &= c\bar{k} = c(w-k). \end{aligned} \tag{22.1.2}$$

The other multiplication rules are determined by the skew commutativity relations (4.2.16) coming from the standard involution; one beautiful consequence is the equality

$$ijk = jki = kij = abc.$$

The R-order $\mathcal{O}$ defined by (22.1.2) is called the **even Clifford algebra** $\operatorname{Clf}^0(Q)$ of Q—its algebra structure is obtained by restriction from the even Clifford algebra of Q_F—and the reduced discriminant of $\mathcal{O}$ is $\operatorname{discrd}(\mathcal{O}) = (N)$. At least one of the minors

$$u^2 - 4bc,\ v^2 - 4ac,\ w^2 - 4ab$$

of the Gram matrix of Q in the standard basis is nonzero since Q is nondegenerate, so for example if $w^2 - 4ab \neq 0$ and $\operatorname{char} F \neq 2$, completing the square we find

$$\mathcal{O} \subset B \simeq \left(\frac{w^2 - 4ab, -aN}{F}\right).$$

It is straightforward to show that the isomorphism class of $\mathcal{O}$ is determined by the similarity class of Q (using the even Clifford algebra construction). Therefore, the proof of Main Theorem 22.1.1 amounts to verifying that every quaternion order arises this way up to isomorphism, and that isomorphic quaternion algebras yield similar ternary quadratic forms.

To this end, we define an inverse to the even Clifford algebra construction. Let $\mathcal{O} \subset B$ be a quaternion order over R with reduced discriminant $\operatorname{discrd}(\mathcal{O})$ generated by $N \in R$ nonzero. Recalling 15.6, let

$$(\mathcal{O}^\sharp)^0 = \{\alpha \in \mathcal{O}^\sharp : \operatorname{trd}(\alpha) = 0\}$$

be the trace zero elements in the dual of $\mathcal{O}$ with respect to the reduced trace pairing. Then we associate the ternary quadratic form

$$\begin{aligned} N\operatorname{nrd}^\sharp(\mathcal{O}) \colon (\mathcal{O}^\sharp)^0 &\to R \\ \alpha &\mapsto N\operatorname{nrd}(\alpha); \end{aligned} \tag{22.1.3}$$

explicitly, we have

$$Ni^\sharp = jk - kj,\ Nj^\sharp = ki - ik,\ Nk^\sharp = ij - ji$$

where $1, i, j, k$ is an R-basis of $\mathcal{O}$, so

$$N \operatorname{nrd}^\sharp(x, y, z) = \operatorname{nrd}\bigl(x(jk - kj) + y(ki - ik) + z(ij - ji)\bigr). \tag{22.1.4}$$

It is then a bit of beautiful algebra to verify that $N \operatorname{nrd}^\sharp$ has discriminant N and that (22.1.3) furnishes an inverse to the even Clifford map.

Just as in the case of fields, the translation from quaternion orders to ternary quadratic forms makes the classification problem easier: we replace the potentially complicated notion of finding a lattice closed under multiplication in a quaternion algebra with the simpler notion of choosing coefficients of a quadratic form.

To conclude this introduction, we state a more general bijective result stated in terms of lattices. Let R be a Dedekind domain with $F = \operatorname{Frac} R$, let $Q_F : V \to F$ be a nondegenerate ternary quadratic form. If $M \subseteq V$ is an R-lattice, and $\mathfrak{l} \subseteq F$ is a fractional ideal of R such that $Q(M) \subseteq \mathfrak{l}$, then we have an induced quadratic form $Q : M \to \mathfrak{l}$; we call such a form a **quadratic module in** V. Given a fractional ideal $\mathfrak{a} \subseteq F$, the **twist** by $\mathfrak{a}$ of the quadratic module $Q : M \to \mathfrak{l}$ in V is the quadratic module $\mathfrak{a}M \to \mathfrak{a}^2\mathfrak{l}$. A **twisted similarity** between quadratic modules Q, Q' in V is a similarity between Q and a twist of Q'. From these notions in hand, we have the following theorem (a special case of Main Theorem 22.5.7).

Theorem 22.1.5. *Let R be a Dedekind domain, and let $Q_F : V \to F$ be a nondegenerate ternary quadratic form. Let $B := \operatorname{Clf}^0 V$. Then the even Clifford map yields a discriminant-preserving bijection*

$$\left\{ \begin{array}{c} \text{Quadratic modules in } V \\ \text{up to twisted similarity} \end{array} \right\} \leftrightarrow \left\{ \begin{array}{c} \text{Quaternion orders in } B \\ \text{up to isomorphism} \end{array} \right\}$$

that is functorial with respect to R.

By *functorial with respect to R*, we mean the same thing as in Corollary 5.2.6, but with respect to any homomorphism $R \to S$ of Dedekind domains. In particular, the bijection in Theorem 22.1.5 is compatible with the bijections obtained over localizations of R, including the bijection over F between quaternion algebras and nondegenerate ternary quadratic forms previously obtained. In the language of quadratic forms (Definition 9.7.13), after some additional work (nailing down the difference between similarity and isometry), we conclude: if the ternary quadratic module Q corresponds to the quaternion order $\mathcal{O}$, then there is a bijection

$$\operatorname{Cl} Q \leftrightarrow \operatorname{Typ} \mathcal{O}, \tag{22.1.6}$$

i.e. the type number of a quaternion order is the same as the class number of the corresponding ternary quadratic form.

Remark 22.1.7. If we restrict the correspondence to *primitive* modules $Q : M \to \mathfrak{l}$ (i.e., $Q(M) = \mathfrak{l}$), then we need only remember the underlying lattice M, and on the right-hand side we obtain precisely the *Gorenstein* orders; these orders will be introduced in 24.1.1 and this correspondence is proven in section 24.2.

22.2 Even Clifford algebras

In this section, we construct the even Clifford algebra associated to a quadratic module: see Remark 22.2.14 below for further references. The reader who wants to skip over technicalities at first is encouraged to skip this section and accept 22.3.2 as a definition.

Let R be a noetherian domain with $F = \operatorname{Frac} R$. Let $Q\colon M \to L$ be a quadratic module over R (see section 9.7), so that M is a projective R-module of finite rank and L is an invertible R-module (rank 1). Write $L^\vee := \operatorname{Hom}_R(L, R)$ and $M^{\otimes 0} = R$. (For further reference on tensor algebra, see Matsumura [Mat89, Appendix C] or Curtis–Reiner [CR81, §12].)

22.2.1. Let

$$\operatorname{Ten}^0(M; L) := \bigoplus_{d=0}^{\infty} (M \otimes M \otimes L^\vee)^{\otimes d}.$$

Now $\operatorname{Ten}^0(M; L)$ has a natural tensor multiplication law (rearranging tensors), so $\operatorname{Ten}^0(M; L)$ is a graded R-algebra. Let $I^0(Q)$ be the two-sided ideal of $\operatorname{Ten}^0(M; L)$ defined by

$$I^0(Q) := \langle x \otimes x \otimes g - g(Q(x)) : x \in M, g \in L^\vee \rangle \subseteq \operatorname{Ten}^0(M; L); \tag{22.2.2}$$

note that $Q(x) \in L$ so $g(Q(x)) \in R$. We define the **even Clifford algebra** of Q to be the quotient

$$\operatorname{Clf}^0(Q) = \operatorname{Ten}^0(M; L)/I^0(Q). \tag{22.2.3}$$

Remark 22.2.4. We might try to define

$$\operatorname{Ten}(M; L) := \bigoplus_{d=0}^{\infty} M^{\otimes d} \otimes (L^\vee)^{\otimes \lfloor d/2 \rfloor} = R \oplus M \oplus (M \otimes M \otimes L^\vee) \oplus \dots;$$

unfortunately, $\operatorname{Ten}(M; L)$ does not have a natural tensor multiplication law, because there is no natural map $M \otimes M \to M \otimes M \otimes L^\vee$. But see 22.2.16 below for the odd part.

Example 22.2.5. Under the inclusion $R \hookrightarrow F$, we have a natural identification

$$\operatorname{Clf}^0(Q) \otimes_R F \cong \operatorname{Clf}^0(Q_F). \tag{22.2.6}$$

We conclude that the R-lattice in $\operatorname{Clf}^0(Q_F)$ defined by the image of R^3 is closed under multiplication—something that may also be verified directly—and so $\operatorname{Clf}^0(Q)$ is an R-order in $\operatorname{Clf}^0(Q_F)$.

22.2.7. As in 5.3.7, for all $x, y \in M$ and $g \in L^\vee$, the calculation

$$x \otimes y \otimes g + y \otimes x \otimes g = g(T(x, y)) \in R \tag{22.2.8}$$

holds in $\operatorname{Clf}^0(Q)$, where $T(x, y) = Q(x + y) - Q(x) - Q(y) \in L$.

22.2.9. If $M \simeq R^n$ is free with basis $e_1, \dots, e_n$ and $L = Rg$ is free, then $\mathrm{Clf}^0(Q)$ is a free R-module with basis

$$e_{i_1} \otimes \cdots \otimes e_{i_d} \otimes g^{-\otimes d/2}, \quad 1 \le e_{i_1} < \cdots < e_{i_d} \le n, \quad d \text{ even},$$

as a consequence of 22.2.8, just as in the case over fields 5.3.9. In particular, by localizing, if M has rank n as an R-module, then $\mathrm{Clf}^0(Q)$ is projective of rank 2^{n-1} as an R-module. We write elements of $\mathrm{Clf}^0(Q)$ without tensors, for brevity.

22.2.10. The **reversal map** defined on simple tensors

$$\begin{aligned} \mathrm{rev} : \mathrm{Clf}^0(Q) &\to \mathrm{Clf}^0(Q) \\ x_1 \otimes \cdots \otimes x_d \otimes (g_1 \cdots g_{d/2}) &\mapsto x_d \otimes \cdots \otimes x_1 \otimes (g_1 \cdots g_{d/2}) \end{aligned}$$

for $x_i \in M$ and $g_i \in L^\vee$, and extended R-linearly, is an R-linear involution.

Theorem 22.2.11. *The association $Q \mapsto \mathrm{Clf}^0(Q)$ is a functor from the category of quadratic R-modules under similarities to the category of projective R-algebras with involution under isomorphism. Moreover, this association is functorial with respect to R.*

We call the association $Q \mapsto \mathrm{Clf}^0(Q)$ in Theorem 22.2.11 the **even Clifford functor**.

22.2.12. The statement "functorial with respect to R" means the following: given a ring homomorphism $R \to S$, there is a natural transformation between the even Clifford functors over R and S. Explicitly, given a ring homomorphism $R \to S$ and a quadratic module $Q\colon M \to L$, we have a quadratic module $Q_S\colon M \otimes_R S \to L \otimes_R S$, and $\mathrm{Clf}^0(Q) \otimes_R S \cong \mathrm{Clf}^0(Q_S)$ in a way compatible with morphisms in each category. In particular, this recovers the identification in Example 22.2.5 arising from $R \hookrightarrow F$.

Remark 22.2.13. The association $Q \mapsto \mathrm{Clf}(Q)$ of the full Clifford algebra is a functor from the category of quadratic R-modules under isometries to the category of R-algebras with involution under isomorphism that is also functorial with respect to R. See Bischel–Knus [BK94].

Proof of Theorem 22.2.11. The construction in 22.2.1 yields an R-algebra that is projective as an R-module; we need to define an association on the level of morphisms. Let $Q' : M' \to L'$ be a quadratic module and (f, h) be a similarity with $f\colon M \xrightarrow{\sim} M'$ and $h : L \xrightarrow{\sim} L'$ satisfying $Q'(f(x)) = h(Q(x))$. We mimic the proof of Lemma 5.3.21. We define a map via

$$\begin{aligned} \mathrm{Ten}^0(M; L) &\to \mathrm{Ten}^0(M'; L') \\ x \otimes y \otimes g &\mapsto f(x) \otimes f(y) \otimes (h^{-1})^*(g) \end{aligned}$$

for $x, y \in M$ and $g \in L^\vee$ and extending multiplicatively, where

$$(h^{-1})^*(g) := g \circ h^{-1} : L' \to R$$

is the pullback under h^{-1}. Then

$$x \otimes x \otimes g - g(Q(x)) \mapsto f(x) \otimes f(x) \otimes (h^{-1})^*(g) - g(Q(x))$$

and since

$$g(Q(x)) = g(h^{-1}(Q'(f(x)))) = (h^{-1})^*(g)(Q'(f(x))),$$

we conclude that $I^0(Q)$ is mapped to $I^0(Q')$. Repeating with the inverse similarity (f^{-1}, h^{-1}), and composing to get the identity, we conclude that the induced map $\mathrm{Clf}^0(Q) \to \mathrm{Clf}^0(Q')$ is an R-algebra isomorphism.

Functoriality in the sense of 22.2.12 then follows directly. □

Remark 22.2.14. In his thesis, Bichsel [Bic85] constructed an even Clifford algebra of a line bundle-valued quadratic form on an affine scheme using faithfully flat descent. A related and more general construction was given by Bischel–Knus [BK94]. Several other constructions are available: see Auel [Auel2011, §1.8] and the references therein.

The direct tensorial construction given above is given for ternary quadratic modules by Voight [Voi2011a, (1.10)] and in general by Auel [Auel2011, §1.8] and with further detail in Auel [Auel2015, §1.2]; for a comparison of this direct construction with others, see Auel–Bernardara–Bolognesi [ABB2014, §1.5, Appendix A].

Remark 22.2.15. Allowing the quadratic forms to take values in a invertible module is essential for what follows and for many other purposes: for an overview, see the introduction to Auel [Auel2011].

22.2.16. Let

$$\mathrm{Ten}^1(M; L) := \bigoplus_{d=1\, d \text{ odd}}^{\infty} M^{\otimes d} \otimes (L^\vee)^{\otimes \lfloor d/2 \rfloor} = M \oplus (M \otimes M \otimes M \otimes L^\vee) \oplus \dots.$$

Then $\mathrm{Ten}^1(M; L)$ is a graded $\mathrm{Ten}^0(M; L)$-bimodule under the natural tensor multiplication. Let $I^1(Q)$ be the R-submodule of $\mathrm{Ten}^1(M; L)$ generated by the image of multiplication of $I^0(Q)$ by M on the left and right: then $I^1(Q)$ is the $\mathrm{Ten}^0(M; L)$-bisubmodule generated by the set of elements of the form

$$x \otimes x \otimes y \otimes g - g(Q(x))y \quad \text{and} \quad y \otimes x \otimes x \otimes g - g(Q(x))y$$

with $x, y \in M$ and $g \in L^\vee$.

We define the **odd Clifford bimodule** as

$$\mathrm{Clf}^1(Q) := \mathrm{Ten}^1(M; L)/I^1(Q).$$

Visibly, $\mathrm{Clf}^1(Q)$ is a bimodule for the even Clifford algebra $\mathrm{Clf}^0(Q)$.

22.2.17. When $L = R$, we can combine the construction of the even Clifford algebra and its odd Clifford bimodule to construct a full Clifford algebra, just as in section 5.3 over a field: see Exercise 22.7. This direct tensorial construction does not extend in an obvious way when $L \neq R$, as we would need to define a multiplication map $M \otimes M \to M \otimes M \otimes L^\vee$.

22.2.18. We will employ exterior calculus in what follows: this is a convenient method for keeping track of our module maps in a general setting. Let M be an R-module and let $r \geq 1$. The r**th exterior power** of M (over R) is

$$\textstyle\bigwedge^r M := M^{\otimes r}/E_r$$

where E_r is the R-module

$$E_r := \langle x_1 \otimes \cdots \otimes x_r : x_1, \dots, x_r \in M \text{ and } x_i = x_j \text{ for some } i \neq j \rangle.$$

We let $\bigwedge^0 M = R$ (and $\bigwedge^1 M = M$). The image of $x_1 \otimes \cdots \otimes x_r \in M^{\otimes r}$ in $\bigwedge^r M$ is written $x_1 \wedge \cdots \wedge x_r$. If M is projective of rank n over R, then $\bigwedge^r M$ is projective of rank $\binom{n}{r}$.

22.3 Even Clifford algebra of a ternary quadratic module

Now suppose that $Q \colon M \to L$ is a ternary quadratic module, which is to say M has rank 3; in this section, we examine its even Clifford algebra $\mathrm{Clf}^0(Q)$. Recall that an R-order is projective if it is projective as an R-module. The main result of this section is as follows.

Theorem 22.3.1. *Let R be a noetherian domain. Then the association $Q \mapsto \mathrm{Clf}^0(Q)$ gives a functor from the category of*

nondegenerateternary quadratic modules over$R$,
under similarities

to the category of

projective quaternion orders over R, under isomorphisms.

In the previous section, we defined the even Clifford functor, whose codomain was the category of projective R-algebras; in this section, we show that the restriction to nondegenerate ternary quadratic modules lands in projective quaternion orders.

We begin with some explicit descriptions.

22.3.2. By 22.2.9, the even Clifford algebra $\mathrm{Clf}^0(Q)$ is an R-algebra that is projective of rank 4 as an R-module. Explicitly, as an R-module we have

$$\mathrm{Clf}^0(Q) \simeq \frac{R \oplus \left(M \otimes M \otimes L^\vee\right)}{I^0(Q)} \tag{22.3.3}$$

where $I^0(Q)$ is the R-submodule generated by elements of the form

$$x \otimes x \otimes g - 1 \otimes g(Q(x))$$

for $x \in M$ and $g \in L^\vee$.

We now explicitly give the even Clifford algebra of a ternary quadratic module in the free case; this could also be taken as the definition when R is a PID and $M = R^3$.

22.3.4. Let $M = R^3$ with standard basis e_1, e_2, e_3 be equipped with the quadratic form $Q\colon M \to R$ defined by

$$Q(x,y,z) = Q(xe_1 + ye_2 + ze_3) = ax^2 + by^2 + cz^2 + uyz + vxz + wxy, \quad (22.3.5)$$

with $a,b,c,u,v,w \in R$. Then

$$N := \operatorname{disc}(Q) = 4abc + uvw - au^2 - bv^2 - cw^2 \in R/R^\times. \quad (22.3.6)$$

By 22.2.9, we have

$$\operatorname{Clf}^0(Q) = R \oplus Ri \oplus Rj \oplus Rk$$

where

$$i := e_2e_3, \quad j := e_3e_1, \quad k := e_1e_2.$$

The reversal involution acts by

$$\bar{i} = e_3e_2 = T(e_2,e_3) - i = u - i,$$

and similarly $\bar{j} = v - j$ and $\bar{k} = w - k$ by (22.2.8).

We then compute directly the multiplication table:

$$\begin{aligned} i^2 &= ui - bc & jk &= a\bar{i} \\ j^2 &= vj - ac & ki &= b\bar{j} \\ k^2 &= wk - ab & ij &= c\bar{k} \end{aligned} \quad (22.3.7)$$

For example,

$$i^2 = (e_2e_3)(e_2e_3) = e_2(e_3e_2)e_3 = e_2(u - e_2e_3)e_3 = ue_2e_3 - e_2^2e_3^2 = ui - bc$$

and

$$jk = (e_3e_1)(e_1e_2) = ae_3e_2 = a\bar{i}.$$

The remaining multiplication laws can be computed in the same way, or by using the reversal involution and (22.3.7): we compute

$$a\bar{i} = \overline{jk} = \bar{k}\,\bar{j} = (w-k)(v-j) = vw - wj - vk + kj$$

so $kj = -vw + ai + wj + vk$. By symmetry, we find:

$$\begin{aligned} kj &= -vw + ai + wj + vk \\ ik &= -uw + wi + bj + uk \\ ji &= -uv + vi + uj + ck \end{aligned} \quad (22.3.8)$$

We note also the formulas

$$ijk = jki = kij = abc. \quad (22.3.9)$$

Example 22.3.10. It is clarifying to work out the diagonal case. Let $B = (a, b \mid F)$ with $a, b \in R$, and let

$$\mathcal{O} = R\langle i, j\rangle = R + Ri + Rj + Rij \subset B$$

be the R-order generated by the standard generators, and let $k = ij$. Then:

$$\begin{aligned} i^2 &= a & jk &= b\bar{i} = -bi \\ j^2 &= b & ki &= a\bar{j} = -aj \\ k^2 &= -ab & ij &= -\bar{k} = k. \end{aligned} \tag{22.3.11}$$

Example 22.3.12. Consider

$$Q(x, y, z) = xy - z^2$$

so $(a, b, c, u, v, w) = (0, 0, -1, 0, 0, 1)$; then $\operatorname{disc}(Q) = -cw^2 = 1$. Then the even Clifford algebra $\operatorname{Clf}^0(Q) = R + Ri + Rj + Rk$ has multiplication table

$$\begin{aligned} i^2 &= 0 & jk &= 0 \\ j^2 &= 0 & ki &= 0 \\ k^2 &= k & ij &= -\bar{k} = k - 1. \end{aligned} \tag{22.3.13}$$

We find an isomorphism of R-algebras

$$\begin{aligned} \operatorname{Clf}^0(Q) &\xrightarrow{\sim} M_2(R) \\ i, j, k &\mapsto \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}, \begin{pmatrix} 0 & 0 \\ -1 & 0 \end{pmatrix}, \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix}. \end{aligned} \tag{22.3.14}$$

22.3.15. Returning to the free quadratic form 22.3.4, the group $\mathrm{GL}_3(R)$ acts naturally on M by change of basis, and this induces an action on $\operatorname{Clf}^0(Q)$ by R-algebra automorphism by functoriality. Explicitly, for $\rho \in \mathrm{GL}_3(R)$, the action on the basis i, j, k is by the *adjugate* $\operatorname{adj}(\rho)$ of ρ, the 3×3 matrix whose entries are the 2×2 minors of ρ. The verification is requested in Exercise 22.2.

22.3.16. Let $F = \operatorname{Frac} R$. By base extension, we have a quadratic form $Q_F : V \to F$ where $V = M \otimes_R F$, and by functoriality 22.2.12 with respect to the inclusion $R \hookrightarrow F$, we have an inclusion $\operatorname{Clf}^0(Q) \hookrightarrow \operatorname{Clf}^0(Q_F)$ realizing $\operatorname{Clf}^0(Q)$ as an R-order in the F-algebra $\operatorname{Clf}^0(Q_F)$.

Lemma 22.3.17. *The reversal involution is a standard involution on* $\operatorname{Clf}^0(Q_F)$.

Proof. To check that the involution is standard, we could appeal to Exercise 3.19, but we find it more illustrative to exhibit the involution on a universal element, yielding a rather beautiful formula. We choose a basis for V and work with the presentation for $\operatorname{Clf}^0(Q_F)$ as in 22.3.4.

Let $\alpha = t + xi + yj + zij$ with $t, x, y, z \in F$. Then $\overline{\alpha} = 2t + ux + vy + wz - \alpha$, and we find that

$$\alpha^2 - (\alpha + \overline{\alpha})\alpha + \alpha\overline{\alpha} = \alpha^2 - \tau(\alpha)\alpha + \nu(\alpha) = 0$$

where

$$\begin{aligned}\tau(\alpha) &= 2t + ux + vy + wz \\ \nu(\alpha) &= t^2 + utx + vty + wtz \\ &\qquad + bcx^2 + (uv - cw)xy + (uw - bv)xz \\ &\qquad + acy^2 + (vw - au)yz + abz^2\end{aligned} \tag{22.3.18}$$

so that the reversal map $\alpha \mapsto \tau(\alpha) - \alpha$ defines a standard involution. □

Lemma 22.3.19. *We have*

$$\operatorname{discrd}(\operatorname{Clf}^0(Q)) = \operatorname{disc}(Q)R.$$

Proof. The construction of the even Clifford algebra is functorial with respect to localization, and the statement itself is local, so we may suppose that $M = R^3$, $L = R$ are free with the presentation for $O = \operatorname{Clf}^0(Q)$ as in 22.3.4.

We refer to section 15.4 and Lemma 15.4.7: we compute

$$\begin{aligned} m(i,j,k) &= \operatorname{trd}((ij - ji)\overline{k}) \\ &= \operatorname{trd}(-2abc + au^2 + cw^2 - aui + (bv - uw)j - cwk) \\ &= -4abc + au^2 + cw^2 - uvw + bv^2 = -\operatorname{disc}(Q) \end{aligned} \tag{22.3.20}$$

and $\operatorname{discrd}(O) = m(i,j,k)R$ as claimed.

Alternatively, we compute directly that

$$\begin{aligned} d(1,i,j,k) &= \begin{pmatrix} 2 & u & v & w \\ u & u^2 - 2bc & cw & bv \\ v & cw & v^2 - 2ac & au \\ w & bv & au & w^2 - 2ab \end{pmatrix} \\ &= -(4abc + uvw - au^2 - bv^2 - cw^2)^2 = -\operatorname{disc}(Q)^2 \end{aligned} \tag{22.3.21}$$

so $\operatorname{disc}(O) = \operatorname{disc}(Q)^2 R$, and the result follows by taking square roots (as ideals). □

Corollary 22.3.22. *If Q is nondegenerate, then* $\operatorname{Clf}^0(Q)$ *is an R-order in the quaternion algebra $B = \operatorname{Clf}^0(Q_F)$.*

Proof. The standard involution has discriminant $\operatorname{disc}(\operatorname{nrd}) = \operatorname{disc}(O)^2 = \operatorname{disc}(Q_F) \neq 0$; the result then follows from the characterization of algebras with nondegenerate standard involution (Main Theorem 4.4.1 and Theorem 6.4.1). □

Remark 22.3.23. Corollary 22.3.22 gives a characteristic independent proof of the fact that the even Clifford algebra of a nondegenerate ternary quadratic form over F is a quaternion algebra over F: we proved this in 5.3.23 and Exercise 6.10 (when $\operatorname{char} F = 2$).

Intermediate between the general abstract definition and the explicit description in the free case is the situation where the modules are completely decomposable, and we can work with a pseudobasis.

Example 22.3.24. Let R be a Dedekind domain. Then we can write

$$M = \mathfrak{a}e_1 \oplus \mathfrak{b}e_2 \oplus \mathfrak{c}e_3 \quad \text{and} \quad L = \mathfrak{l}$$

for fractional ideals $\mathfrak{a}, \mathfrak{b}, \mathfrak{c}, \mathfrak{l}$. Let $V = M \otimes_R F \simeq F^3$ with basis e_1, e_2, e_3, so then $M \hookrightarrow V$ is a ternary R-lattice. Then we may take $Q_F : V \to F$ to have the form 22.3.5, and $\mathrm{Clf}^0(Q_F) = B$ is a quaternion algebra with $O := \mathrm{Clf}^0(Q) \subseteq B$ an R-order.

Extending the description in 22.3.4, we find that

$$O = R \oplus \mathfrak{b}\mathfrak{c}\mathfrak{l}^{-1} i \oplus \mathfrak{a}\mathfrak{c}\mathfrak{l}^{-1} j \oplus \mathfrak{a}\mathfrak{b}\mathfrak{l}^{-1} k \tag{22.3.25}$$

where i, j, k satisfy the multiplication table (22.3.7). We can verify directly that O is closed under multiplication: for example, if $\alpha \in \mathfrak{b}\mathfrak{c}\mathfrak{l}^{-1}$ so $\alpha i \in O$, then

$$(\alpha i)^2 = u\alpha i - \alpha^2 bc \in O$$

since $Q(\mathfrak{b}e_2) = \mathfrak{b}^2 Q(e_2) \subseteq \mathfrak{l}$ so $b = Q(e_2) \in \mathfrak{l}\mathfrak{b}^{-2}$ and therefore

$$\alpha^2 bc \in (\mathfrak{b}\mathfrak{c}\mathfrak{l}^{-1})^2 (\mathfrak{l}\mathfrak{b}^{-2})(\mathfrak{l}\mathfrak{c}^{-2}) = R.$$

Example 22.3.26. Let $F = \mathbb{Q}(\sqrt{10})$ and $R = \mathbb{Z}_F = \mathbb{Q}[\sqrt{10}]$ be the ring of integers. Then $\mathfrak{p} = (3, 4 + \sqrt{10})$ is a prime ideal over 3 that is not principal.

Let $Q : M = R^3 \to \mathfrak{p}$ be the quadratic module

$$Q(x, y, z) = 3x^2 + 3y^2 + (4 + \sqrt{10})z^2.$$

We have $\mathfrak{p} = Q(R^3)$. The even Clifford algebra is then

$$O = \mathrm{Clf}^0(Q) = R + \mathfrak{p}^{-1} i + \mathfrak{p}^{-1} j \oplus \mathfrak{p}^{-1} k$$

with the multiplication law

$$\begin{aligned} i^2 &= -3(4 + \sqrt{10}) & jk &= 3\bar{i} \\ j^2 &= -3(4 + \sqrt{10}) & ki &= 3\bar{j} \\ k^2 &= -9 & ij &= (4 + \sqrt{10})\bar{k}. \end{aligned} \tag{22.3.27}$$

We have

$$\mathrm{discrd}(O) = 4(9)(4 + \sqrt{10})\mathfrak{p}^{-3} = (2, \sqrt{10})^5 (3, 2 + \sqrt{10})^2$$

and in particular $\mathfrak{p} \nmid \mathrm{discrd}(O)$, and

$$O \subset B = \left(\frac{-3(4 + \sqrt{10}), -3(4 + \sqrt{10})}{F} \right)$$

with $\mathrm{disc}\, B = (2 + w)R$, so $\mathrm{Ram}\, B = \{(2, \sqrt{10}), \mathfrak{p}, \infty_1, \infty_2\}$ where ∞_1, ∞_2 are the two real places of F.

22.4 Over a PID

In the previous two sections, we observed that the construction of the even Clifford algebra gives a functorial association from nondegenerate ternary quadratic modules to quaternion orders. In this section, we show that this functor gives a bijection on classes over a PID, following Gross–Lucianovic [GrLu2009, §4].

Main Theorem 22.4.1. *Suppose that R is a PID. Then the association $Q \mapsto \mathrm{Clf}^0(Q)$ induces a discriminant-preserving bijection*

$$\left\{ \begin{array}{c} \text{Nondegenerate ternary quadratic} \\ \text{forms over } R \text{ up to similarity} \end{array} \right\} \leftrightarrow \left\{ \begin{array}{c} \text{Quaternion orders over } R \\ \text{up to isomorphism} \end{array} \right\} \tag{22.4.2}$$

that is functorial with respect to R.

Remark 22.4.3. The bijection can also be rephrased in terms of the orbits of a group (following Gross–Lucianovic [GrLu2009]). The group $\mathrm{GL}_3(R)$ has a natural twisted action on quadratic forms by $(gQ)(x,y,z) = (\det g)(Q(g^{-1}(x,y,z)^{\mathsf{t}}))$, i.e., the usual action with an extra scaling factor of $\det g \in R^\times$. This is the natural action on the R-module $\mathrm{Sym}^2((R^3)^\vee) \otimes \bigwedge^3 R^3$, or equivalently on the set of quadratic modules $Q\colon R^3 \to \bigwedge^3 R^3$. Main Theorem 22.1.1 states that the nondegenerate orbits of this action are in functorial bijection with the set of isomorphism classes of quaternion orders over R.

We prove this theorem in a few steps. Throughout this section, let R be a PID.

First, we prove that the map (22.4.2) is surjective, or equivalently that the even Clifford functor is essentially surjective from the category of nondegenerate ternary quadratic forms to the category of quaternion orders.

Proposition 22.4.4. *Every quaternion R-order is isomorphic to the even Clifford algebra of a nondegenerate ternary quadratic form.*

Proof. We work explicitly with the multiplication table, hoping to make it look like (22.3.7).

Let $\mathcal{O}$ be a quaternion R-order. Since R is a PID, $\mathcal{O}$ is free as an R-module. We need a slight upgrade from this, a technical result supplied by Exercise 22.1: in fact, $\mathcal{O}$ has an R-basis containing 1.

So let $1, i, j, k$ be an R-basis for $\mathcal{O}$. Since every element of $\mathcal{O}$ is integral over R, satisfying its reduced characteristic polynomial of degree 2 over R, we have

$$\begin{aligned} i^2 &= ui + l \\ j^2 &= vj + m \\ k^2 &= wk + n \end{aligned}$$

for some $l,m,n,u,v,w \in R$. The product $jk = r - ai + qj + \alpha k$ can be written as an R-linear combination of $1, i, j, k$, with $q, r, a, \alpha \in R$. Letting $k' := k - q$, we have

$$jk' = j(k-q) = r - ai + \alpha k = (r + \alpha q) - ai + \alpha k'.$$

So changing the basis, we may suppose jk is an R-linear combination of $1, i, k$ (no j term). By symmetry, in the product ki we may suppose that the coefficient of k is zero and in ij the coefficient of i is zero. Therefore:

$$\begin{aligned} jk &= r - ai + \alpha k \\ ki &= s - bj + \beta i \\ ij &= t - ck + \gamma j \end{aligned}$$

As before, the other products can be calculated using the standard involution: for example, we have

$$\begin{aligned} ik + ki &= -\operatorname{trd}(k\bar{i}) + \operatorname{trd}(k)i + \operatorname{trd}(i)k \\ &= -\operatorname{trd}(k(u-i)) + wi + uk \\ &= (-uw + 2s - bv + \beta u) + wi + uk \end{aligned}$$

so

$$ik = (s + \beta u - bv - uw) + (w - \beta)i + bj + uk. \tag{22.4.5}$$

But now from these multiplication laws, we compute that the trace of left multiplication i is $\operatorname{Tr}(i) = 0+u+\gamma+u = 2u+\gamma$. But in a quaternion algebra, we have $\operatorname{Tr}(i) = 2\operatorname{trd}(i) = 2u$, so we must have $\gamma = 0$. By symmetry, we find that $\alpha = \beta = 0$. Finally, associativity implies relations on the structure constants in the multiplication table: we have

$$\begin{aligned} j(k\bar{k}) &= (jk)\bar{k} \\ -nj &= (r - ai)(w - k) = rw - awi - rk + aik \\ -nj &= (rw + as - abv - auw) + abj + (au - r)k \end{aligned} \tag{22.4.6}$$

using (22.4.5) with $\beta = 0$; so equality of coefficients of j, k implies $r = au$ and $n = -ab$. By symmetry, we find $s = bv, t = cw$ and $m = -ac, n = -ab$, so we have the following multiplication table:

$$\begin{aligned} i^2 &= ui - bc & \qquad jk &= a\bar{i} \\ j^2 &= vj - ac & ki &= b\bar{j} \\ k^2 &= wk - ab & ij &= c\bar{k} \end{aligned}$$

This matches precisely the multiplication table (22.3.7) for the even Clifford algebra of the quadratic form $Q(x, y, z) = ax^2 + by^2 + cz^2 + uyz + vxz + wxy$. □

22.4.7. More generally, if R is a domain and O is a quaternion R-order such that O is free as an R-module with basis $1, i, j, k$, then the proof of Proposition 22.4.4 shows O has a basis $1, i, j, k$ satisfying the multiplication laws (22.3.7) of an even Clifford algebra; we call such a basis a **good basis** for O. Moreover, we have seen that given a basis $1, i, j, k$, there exist unique $\eta(i), \eta(j), \eta(k) \in R$ (in fact, certain coefficients of the multiplication table) such that

$$1,\ i - \eta(i),\ j - \eta(j),\ k - \eta(k)$$

is a good basis.

To conclude, we need to show that if two quaternion R-orders are isomorphic, then they correspond to similar ternary quadratic forms. To this end, we define an inverse.

22.4.8. Let $\mathcal{O} \subseteq B$ be a quaternion R-order with R-basis $1, i, j, k$. Let $N \in R$ be such that $(N) = \operatorname{discrd}(\mathcal{O})$; then $N \neq 0$ and is well-defined up to multiplication by $R^\times$. Let $1^\sharp, i^\sharp, j^\sharp, k^\sharp$ be the dual basis (see 15.6.3); then $\operatorname{trd}(i^\sharp) = \operatorname{trd}(1 \cdot i^\sharp) = 0$ and similarly for $j^\sharp, k^\sharp$, so

$$(\mathcal{O}^\sharp)^0 = \{\alpha \in \mathcal{O}^\sharp : \operatorname{trd}(\alpha) = 0\} = Ri^\sharp + Rj^\sharp + Rk^\sharp.$$

We define a candidate quadratic form

$$N \operatorname{nrd}^\sharp(\mathcal{O})(x, y, z) = N \operatorname{nrd}(xi^\sharp + yj^\sharp + zk^\sharp), \tag{22.4.9}$$

well-defined up to similarity (along the way, we chose a basis and a generator for $\operatorname{discrd}(\mathcal{O})$).

Example 22.4.10. We return to Example 22.3.10. The R-order $\mathcal{O}$ has reduced discriminant $N = 4ab$. The (rescaled) dual basis is

$$Ni^\sharp = 2bi, \quad Nj^\sharp = 2aj, \quad Nk^\sharp = -2k$$

and $i^\sharp, j^\sharp, k^\sharp$ is a basis for $(\mathcal{O}^\sharp)^0$; thus

$$N \operatorname{nrd}(xi^\sharp + yj^\sharp + zk^\sharp) = \frac{1}{N}(-4ab^2x^2 - 4a^2by^2 + 4abz^2) = -bx^2 - ay^2 + z^2.$$

Example 22.4.11. We return to Example 22.3.12. We have

$$i^\sharp, j^\sharp, k^\sharp = \begin{pmatrix} 0 & 0 \\ 1 & 0 \end{pmatrix}, \begin{pmatrix} 0 & -1 \\ 0 & 0 \end{pmatrix}, \begin{pmatrix} -1 & 0 \\ 0 & 1 \end{pmatrix}$$

and $N = 1$ so

$$N \operatorname{nrd}(xi^\sharp + yj^\sharp + zk^\sharp) = \det \begin{pmatrix} -z & -y \\ x & z \end{pmatrix} = xy - z^2.$$

Proposition 22.4.12. *If $Q\colon R^3 \to R$ is a nondegenerate ternary quadratic form with $\operatorname{disc} Q = N$, then $N \operatorname{nrd}^\sharp(\operatorname{Clf}^0(Q))$ is similar to Q. If $\mathcal{O}$ is a quaternion R-order with $\operatorname{discrd}(\mathcal{O}) = (N)$, then $N \operatorname{nrd}^\sharp(\mathcal{O})\colon R^3 \to R$ has $\operatorname{Clf}^0(N \operatorname{nrd}^\sharp(\mathcal{O})) \simeq \mathcal{O}$.*

Proof. Proposition 22.4.4 shows that the even Clifford functor induces a surjective map from similarity classes of nondegenerate ternary quadratic forms over R to isomorphism classes of quaternion R-orders. If we prove the first statement, then the second follows from set theory (and can be verified in a similar way).

We start with the quadratic form (22.3.5) with $\mathcal{O}$ satisfying the multiplication laws (22.3.7). Let $N := \operatorname{disc}(Q)$. We claim that

$$\begin{aligned} Ni^\sharp &= jk - kj = (au + vw) - 2ai - wj - vk \\ Nj^\sharp &= ki - ik = (bv + uw) - wi - 2bj - uk \\ Nk^\sharp &= ij - ji = (cw + uv) - vi - uj - 2ck. \end{aligned} \tag{22.4.13}$$

We see that $\operatorname{trd}(Ni^\sharp) = 0$ and the same with $j^\sharp, k^\sharp$. We recall the alternating trilinear form m (defined in 15.4.2). By (22.3.20) we have

$$m(i, j, k) = -N = \operatorname{trd}(\bar{i}(jk - kj)) = -\operatorname{trd}(i(jk - kj)) = \operatorname{trd}(i(Ni^\sharp))$$

and

$$m(j, j, k) = 0 = \operatorname{trd}(\bar{j}(jk - kj)) = -\operatorname{trd}(j(Ni^\sharp))$$

and similarly $\operatorname{trd}(k(Ni^\sharp)) = 0$. The other equalities follow similarly, and this verifies the dual basis (22.4.13). In particular, we have $\operatorname{trd}(Ni^\sharp) = \operatorname{trd}(Nj^\sharp) = \operatorname{trd}(Nk^\sharp) = 0$.

We then compute the quadratic form on this basis and claim that

$$\begin{aligned} \operatorname{nrd}(N(xi^\sharp + yj^\sharp + zk^\sharp)) &= N(ax^2 + by^2 + cz^2 + uyz + vxz + wxy) \\ &= NQ(x, y, z). \end{aligned} \tag{22.4.14}$$

Indeed,

$$2\operatorname{nrd}(Ni^\sharp) = \operatorname{trd}((Ni^\sharp)(\overline{Ni^\sharp})) = -\operatorname{trd}((Ni^\sharp)^2) = -(-2a)N = 2aN$$

since only the term $\operatorname{trd}(Ni^\sharp i) = N$ is nonzero; and

$$\operatorname{trd}(Ni^\sharp \overline{Nj^\sharp}) = -\operatorname{trd}(Ni^\sharp Nj^\sharp) = -wN.$$

The other equalities follow by symmetry. Then the claim (22.4.14) implies that $N\operatorname{nrd}(xi^\sharp + yj^\sharp + zk^\sharp) = Q(x, y, z)$, as desired. □

Corollary 22.4.15. *Let $\mathcal{O}$ be a quaternion R-order. Then*

$$\mathcal{O} = R + \operatorname{discrd}(\mathcal{O})(\mathcal{O}^\sharp)^0(\mathcal{O}^\sharp)^0 = R + \operatorname{discrd}(\mathcal{O})\mathcal{O}^\sharp\mathcal{O}^\sharp.$$

Proof. If we take the identifications in the proof of Proposition 22.4.12 working within $B \supseteq \mathcal{O}$, we see that $\operatorname{Clf}^0(N\operatorname{nrd}^\sharp(\mathcal{O}))$ is spanned over R by the elements

$$1, Ni^\sharp j^\sharp, Nj^\sharp k^\sharp, Nk^\sharp i^\sharp$$

where $\operatorname{discrd}(\mathcal{O}) = (N)$. In order to see that the other factors belong to this ring, we compute

$$\begin{aligned} (Ni^\sharp)^2 &= -aN \\ (Nj^\sharp)^2 &= -bN \\ (Nk^\sharp)^2 &= -cN. \end{aligned} \tag{22.4.16}$$

and

$$\begin{aligned} (Nj^\sharp)(Ni^\sharp) &= -N\bar{k} \\ (Nk^\sharp)(Nj^\sharp) &= -N\bar{i} \\ (Ni^\sharp)(Nk^\sharp) &= -N\bar{j}. \end{aligned} \tag{22.4.17}$$

If we want to throw in the factors with $1^\sharp$ as well, then we check:

$$\begin{aligned} N1^\sharp &= 2N - ii^\sharp - jj^\sharp - kk^\sharp \\ &= N - 2(abc + uvw) + (au + vw)i + (bv + uw)j + (cw + uv)k. \end{aligned}$$

satisfies

$$(N1^\sharp)^2 - N(N1^\sharp) + N(abc + uvw) = 0 \tag{22.4.18}$$

and

$$\begin{aligned}
(N1^\sharp)(Ni^\sharp) &= -N(au + vw - ai - vk) = (Ni^\sharp)(N1^\sharp) + N(wj - vk) \\
(N1^\sharp)(Nj^\sharp) &= -N(bv + wu - wi - bj) = (Nj^\sharp)(N1^\sharp) + N(-wi + uk) \\
(N1^\sharp)(Nk^\sharp) &= -N(cw + uv - uj - ck) = (Nk^\sharp)(N1^\sharp) + N(vi - uj).
\end{aligned}$$

The result follows. □

Finally, we officially combine our work to prove the main theorem of this section.

Proof of Main Theorem 22.4.1. Combine Propositions 22.4.4 and 22.4.12. □

Remark 22.4.19. Just as in section 5.5, we may ask about embeddings of a quadratic ring in an order. However, moving from the rational to the integral is a bit tricky, and the issue of embeddings is a theme that will return with gusto in Chapter 30. In that context, it will be more natural to look at a different ternary quadratic form to measure embeddings; just as in the case of trace zero, it is related to but not the same as the one obtained in the above bijection.

22.5 Twisting and final bijection

In this final section, we conclude with the final bijection. We must keep track of the extra data of an ideal class, and along the way allow coefficient ideals. Throughout, let R be a Dedekind domain.

We first need the following slightly revised notion of similarity (one that 'glues together' local similarities) allowing scaling by fractional ideals.

Definition 22.5.1. Let $Q\colon M \to \mathfrak{l}$ be a quadratic module with $\mathfrak{l}$ a fractional R-ideal. The **twist** of Q by a fractional R-ideal $\mathfrak{u}$ is the quadratic form $\mathfrak{u} \otimes Q\colon \mathfrak{u} \otimes M \to \mathfrak{u}^2\mathfrak{l}$ defined by $(\mathfrak{u} \otimes Q)(u \otimes x) = u^2 Q(x)$.

A **twisted similarity** between quadratic modules Q and Q' is a similarity between Q and a twist $\mathfrak{u} \otimes Q'$ for some fractional R-ideal $\mathfrak{u}$.

Example 22.5.2. If $\mathfrak{u} = uR$ is a principal fractional ideal, then twisted similarities between Q and $\mathfrak{a}Q' = uQ'$ are precisely those obtained from a similarity between Q and Q', multiplied by u. In particular, if R is a PID, then the notions of similarity and twisted similarity coincide.

Example 22.5.3. Two quadratic modules $Q, Q'\colon M, M' \to \mathfrak{l}$ with the *same* codomain are twisted similar if and only if they are similar. Indeed, if $\mathfrak{u}^2\mathfrak{l} = \mathfrak{l}$, then $\mathfrak{u} = R$.

22.5.4. Second, we extend the definition of the inverse in 22.4.8 using the reduced norm to the noetherian domain R as follows. Let $\mathcal{O} \subseteq B$ be a quaternion R-order. Since

R is a Dedekind domain, the reduced discriminant $\operatorname{discrd}(O) \subseteq R$ of O is an invertible R-ideal. Define the map

$$\begin{aligned} \operatorname{nrd}^\sharp(O)\colon (O^\sharp)^0 &\to F \\ \alpha &\mapsto \operatorname{nrd}(\alpha). \end{aligned} \tag{22.5.5}$$

Lemma 22.5.6. *For a quaternion order O, the map $\operatorname{nrd}^\sharp(O)$ defines a ternary quadratic module with values in $\operatorname{discrd}(O)^{-1}$.*

Proof. The reduced norm defines a quadratic map, so we only need to verify that the codomain is valid. To this end, we may check locally since reduced norm and reduced discriminant commute with localization. Reducing to the local case, suppose R is now a local Dedekind domain hence a PID. Choosing a basis, we verified in (22.4.14) that $N\operatorname{nrd}^\sharp(O) \subseteq R$, where $\operatorname{discrd}(O) = (N)$; the result follows. □

Main Theorem 22.5.7. *Let R be a Dedekind domain. Then the associations*

$$\begin{gathered} \left\{ \begin{array}{c} \text{Nondegenerate ternary quadratic} \\ \text{modules over } R \\ \text{up to twisted similarity} \end{array} \right\} \leftrightarrow \left\{ \begin{array}{c} \text{Quaternion orders} \\ \text{over } R \text{ up to} \\ \text{isomorphism} \end{array} \right\} \\ Q \mapsto \operatorname{Clf}^0(Q) \\ \operatorname{nrd}^\sharp(O) \leftarrow\!\shortmid O \end{gathered} \tag{22.5.8}$$

are mutually inverse, discriminant-preserving bijections that are also functorial with respect to R.

Proof. We proved a version of this statement when R is a PID in Theorem 22.4.1. More generally, we work now with a pseudobasis instead of a basis, explaining the presence of the twisted similarity.

The surjectivity of the even Clifford map follows by generalizing the argument in Proposition 22.4.4 and 22.4.7 to show that O has a *good pseudobasis*: see Exercise 22.5.

Let $Q\colon M \to \mathfrak{l}$ be a quadratic module with $O := \operatorname{Clf}^0(Q)$. Returning to Example 22.3.24, we may write

$$M = \mathfrak{a}e_1 \oplus \mathfrak{b}e_2 \oplus \mathfrak{c}e_3$$

for fractional ideals $\mathfrak{a}, \mathfrak{b}, \mathfrak{c}$, and with $Q_F(x, y, z) := Q(xe_1 + ye_2 + ze_3)$ in the usual form provided by (22.3.5). Let $N := \operatorname{disc}(Q_F)$. Then

$$O = R \oplus \mathfrak{b}\mathfrak{c}\mathfrak{l}^{-1} i \oplus \mathfrak{a}\mathfrak{c}\mathfrak{l}^{-1} j \oplus \mathfrak{a}\mathfrak{b}\mathfrak{l}^{-1} k.$$

Consider now $\operatorname{nrd}^\sharp(O)\colon (O^\sharp)^0 \to \operatorname{discrd}(O)^{-1}$: as in 22.4.8 we have

$$(O^\sharp)^0 = \mathfrak{l}(\mathfrak{b}\mathfrak{c})^{-1} i^\sharp \oplus \mathfrak{l}(\mathfrak{a}\mathfrak{c})^{-1} j^\sharp \oplus \mathfrak{l}(\mathfrak{a}\mathfrak{b})^{-1} k^\sharp. \tag{22.5.9}$$

To prepare our twisted similarity, let

$$\mathfrak{d} := \mathfrak{a}\mathfrak{b}\mathfrak{c}\mathfrak{l}^{-1} \simeq \textstyle\bigwedge^3 M \otimes L^\vee. \tag{22.5.10}$$

Then

$$\mathfrak{d}(\mathcal{O}^\sharp)^0 = \mathfrak{a}i^\sharp \oplus \mathfrak{b}j^\sharp \oplus \mathfrak{c}k^\sharp. \tag{22.5.11}$$

We claim that the reduced norm on $\mathfrak{d}(\mathcal{O}^\sharp)^0$ has values in $N^{-1}\mathfrak{l}$ and is similar to Q. The claim follows from the same calculation in the proof of Proposition 22.4.12, namely that $\operatorname{nrd}(xi^\sharp + yj^\sharp + zk^\sharp) = N^{-1}Q_F(x, y, z)$! We conclude that $\operatorname{nrd}^\sharp(\mathcal{O})$ is twisted similar to Q. □

We conclude with the following application to quadratic forms.

Corollary 22.5.12. *Let $Q: M \to \mathfrak{l}$ be a ternary quadratic module and $\mathcal{O} := \operatorname{Clf}^0(Q)$. Then the even Clifford map induces a bijection $\operatorname{Cl} Q \leftrightarrow \operatorname{Typ} \mathcal{O}$.*

Proof. We first claim that the even Clifford map induces an injection $\operatorname{Gen} Q \to \operatorname{Gen} \mathcal{O}$, giving an injection $\operatorname{Cl} Q \to \operatorname{Typ} \mathcal{O}$. Indeed, let $Q' \in \operatorname{Gen} Q$, so $Q': M' \to \mathfrak{l}'$ is locally isometric to Q. Let $\mathcal{O}' := \operatorname{Clf}^0(Q')$. Since Q', Q are locally isometric, they are locally similar, so $\mathcal{O}', \mathcal{O}$ are locally isomorphic by Main Theorem 22.4.1, so $\mathcal{O}' \in \operatorname{Gen} \mathcal{O}$. And if $Q' \simeq Q$ are isometric, again they are (twisted) similar, so by Main Theorem 22.5.7 we have $\mathcal{O}' \simeq \mathcal{O}$.

To finish, we need to show that the even Clifford map is surjective. We pass from similarity classes to isometry classes in the same way as in the proof of Corollary 5.2.6. To this end, let $\mathcal{O}' \in \operatorname{Gen}(\mathcal{O})$. Let $Q' := \mathfrak{d} \operatorname{nrd}^\sharp(\mathcal{O}')$ as in (22.5.11). By the same rescaling argument given in Corollary 5.2.6, applying a similarity to Q' we may further suppose that $\operatorname{disc} Q'_F = \operatorname{disc} Q_F \in F^\times/F^{\times 2}$. By surjectivity in Main Theorem 22.5.7, for every prime $\mathfrak{p}$ of R, there exists a twisted similarity from $Q'_{(\mathfrak{p})}$ to $Q_{(\mathfrak{p})}$ over $R_{(\mathfrak{p})}$—and since each $R_{(\mathfrak{p})}$ is a PID, by Example 22.5.2, these are in fact similarities. Taking such a similarity and considering it as a similarity over F, again repeating the same argument as at the end of Corollary 5.2.6, we conclude that Q' and Q are locally isometric, so $Q' \in \operatorname{Gen} Q$. Finally, if $\mathcal{O}' \simeq \mathcal{O}$, repeating these arguments one more time over R (first to go from twisted similar to similar, then to note the similarity gives rise to an isometry) we conclude that $Q' \simeq Q$. □

Remark 22.5.13. The correspondence between ternary quadratic forms and quaternion orders has a particularly rich history. Perhaps the earliest prototype is due to Hermite [Herm1854], who examined the product of automorphs of ternary quadratic forms. Early versions of the correspondence were given by Latimer [Lat37, Theorem 3], Pall [Pall46, Theorems 4–5], and Brandt [Bra43, §3ff] over $\mathbb{Z}$ by use of explicit formulas.

Various attempts were made to generalize the correspondence to Dedekind domains, with the thorny issue being how to deal with a nontrivial class group. Eichler [Eic53, §14, p. 96] gave such an extension. Peters [Pet69, §4] noted that Eichler's correspondence was not onto due to class group issues, and he gave a rescaled version that gives a bijection for Gorenstein orders. Eichler's correspondence was further tweaked by Nipp [Nip74, §3], who opted for a different scaling factor that is not restricted to a class of orders, but his correspondence fails to be onto [Nip74, p.536].

These correspondences were developed further by Brzezinski [Brz80, §3], [Brz85, §3], where he connected the structure of orders to relatively minimal models of the corresponding integral conic; see also Remark 24.3.11. He revisited the correspondence

again in the context of Gorenstein orders [Brz82, §3] and Bass orders [Brz83b, §2]. Lemurell [Lem2011, Theorem 4.3] gives a concise account of the correspondence of Brzezinski over a PID (the guts of which are contained in [Brz82, (3.2)]).

More recently, Gross–Lucianovic [GrLu2009, §4] revisited the correspondence over a PID or local ring, and they extended it to include quadratic forms of nonzero discriminant and without restricting to Gorenstein orders; this extension is important for automorphic reasons, connected to Fourier coefficients of modular forms on PGSp(6), as developed by Lucianovic in his thesis [Luc2003]. Balaji [Bal2007, Theorem 3.1] studied degenerations of ternary quadratic modules in the context of orthogonal groups and Witt invariants and showed that the even Clifford functor is bijective over a general scheme. Finally, Voight [Voi2011a, Theorem B] gave a general and functorial correspondence without any of the above restrictive hypotheses, including the functorial inverse to the even Clifford functor provided above.

Remark 22.5.14. In the most general formulation of the correspondence, allowing arbitrary ternary quadratic modules discriminant over all sorts of rings, Voight [Voi2011a, Theorem A] characterizes the image of the even Clifford functor, as follows. Let B be an R-algebra that is (faithfully) projective of rank 4 as an R-module. Then B is a **quaternion ring** if $B \simeq \operatorname{Clf}^0(Q)$ for a ternary quadratic module Q. Then B is a quaternion ring if and only if B has a standard involution and for all $x \in B$, the trace of left (or right) multiplication by x on B is equal to $2\operatorname{trd}(x)$.

For example, if we take the quadratic form $Q\colon R^3 \to R$ defined by $Q(x,y,z) = 0$ identically, the multiplication table on $\operatorname{Clf}^0(Q)$ gives the commutative ring

$$\operatorname{Clf}^0(Q) \simeq R[i,j,k]/(i,j,k)^2.$$

One can see this as a kind of deformation of a quaternion algebra (in an algebro-geometric sense), letting $a, b \to 0$.

Exercises

▹ 1. Let R be a PID or local noetherian domain. Let A be an R-algebra that is free of finite rank as an R-module. Show that A has an R-basis including 1. *[Hint: show that the quotient A/R is torsion-free, hence free; since free modules are projective, the sequence $0 \to R \to A \to A/R \to 0$ splits, giving $A \simeq R \oplus A/R$.]*

▹ 2. For a free quadratic ternary form (as in 22.3.4), show that a change of basis $\rho \in \mathrm{GL}_3(R)$ acts on $i, j, k \in \operatorname{Clf}^0(Q)$ by the adjugate matrix $\operatorname{adj}(\rho) \in \mathrm{GL}_3(R)$ (where the entries of $\operatorname{adj}(\rho)$ are the 2×2-cofactors of ρ and $\rho\operatorname{adj}(\rho) = \det(\rho)$).

3. Let R be a domain and let $\operatorname{Pic} R$ be the group of isomorphism classes of invertible R-modules (equivalently, classes of fractional R-ideals in F). Show that up to twisted similarity, the target of a quadratic module only depends on its class in $\operatorname{Pic} R/2\operatorname{Pic} R$. [See Example 9.7.5.]

4. Finish the direct verification in Example 22.3.24 that O is closed under multiplication.

▹ 5. Let R be a Dedekind domain, and let O be a quaternion R-order. Show that there exist $i, j, k \in O$ and $\mathfrak{a}, \mathfrak{b}, \mathfrak{c} \subset F$ fractional R-ideals such that $O = R + \mathfrak{a}i + \mathfrak{b}j + \mathfrak{c}k$ and such that $1, i, j, k$ satisfy the multiplication rules (22.3.7) for some $a, b, c, u, v, w \in$

F—called a **good pseudobasis** for $\mathcal{O}$. *[Hint: revisit what goes into 22.4.7; a simple observation will suffice!]*

6. Let $Q: R^3 \to R$ be a ternary quadratic form and let $\mathcal{O} := \mathrm{Clf}^0(Q)$. Show that
$$\operatorname{disc} \operatorname{nrd}|_{\mathcal{O}} = (\operatorname{disc} Q)^2.$$
[Hint: see (22.3.18).]

7. Let $Q: M \to R$ be a quadratic form over R. Construct a Clifford algebra with a universal property analogous to Proposition 5.3.1, and recover the even Clifford algebra and odd Clifford bimodule.

8. Let $Q: M \to R$ be a quadratic form such that there exists $x \in M$ such that $Q(x) \in R^\times$.

   (a) Show that the odd Clifford bimodule $\mathrm{Clf}^1(Q)$ is free of rank 1 as a $\mathrm{Clf}^0(Q)$.
   (b) Generalize this result to case where $Q: M \to L$ is a quadratic module.

9. Let $Q: M \to R$ be a quadratic form over R with M of odd rank as an R-module and let $F = \operatorname{Frac} R$. Let $S := Z(\mathrm{Clf}\, Q) \hookrightarrow K := Z(\mathrm{Clf}\, Q_F)$ be the center of the Clifford algebra of Q. Show that S is an R-order in K.

▹ 10. Show that $\operatorname{nrd}(\mathcal{O}^\sharp) = \operatorname{nrd}((\mathcal{O}^\sharp)^0)$. *[Hint: use* (22.4.18)*.]*

11. Let R be a Dedekind domain with $F = \operatorname{Frac} R$, let B be a quaternion algebra over F, and let $\mathcal{O} \subseteq B$ be an R-order. Let $S \subseteq \mathcal{O}$ be an R-order.

   (a) Suppose $S \subseteq \mathcal{O}$ is integrally closed. Prove that $\mathcal{O}$ is projective of rank 2 as a left S-module.
   (b) If S is not integrally closed, then show that (a) need not hold by the following example. Let $R = \mathbb{Z}$ and $F = \mathbb{Q}$, let $B = (-1, -1 \mid \mathbb{Q})$, let
$$\mathcal{O} = \mathbb{Z} + \mathbb{Z}pi + \mathbb{Z}j + \mathbb{Z}ij$$
for an odd prime p (that is pi, not π!). Let $S = \mathbb{Z}[pi] \subseteq \mathcal{O}$. Show that $\mathcal{O}$ is not projective as a left S-module.
   (c) Show that the property that $\mathcal{O}$ is projective as an S-module is a local property (over primes of R).
   (d) In light of (c), suppose that R is a PID, and write $\mathcal{O}$ in a good basis (22.3.7). Suppose that $S = R[i]$ with $i^2 = ui - bc$. Show that $\mathcal{O}$ is projective as an S-module if and only if the quadratic form $bx^2 + uxy + cy^2$ represents a unit.
   (e) Using (d), conclude in general that if S has conductor coprime to $\operatorname{discrd} \mathcal{O}$, show that $\mathcal{O}$ is projective as an S-module.

12. Let R be a global ring with $F := \operatorname{Frac} R$, let B, B' be quaternion algebras over F, and let $\mathcal{O} \subseteq B$ and $\mathcal{O}' \subseteq B'$ be R-orders. Consider the *quaternary* quadratic forms $Q := \operatorname{nrd}|_{\mathcal{O}}: \mathcal{O} \to R$ and similarly Q' on $\mathcal{O}'$.

   (a) Show that Q is isometric to Q', then there is an isomorphism of F-algebras $B \xrightarrow{\sim} B'$.

(b) In light of (a), suppose $\mathcal{O}, \mathcal{O}' \subseteq B = B'$. Show that $\mathcal{O}, \mathcal{O}'$ are isomorphic as R-orders if and only if Q, Q' are isometric as (quaternary) quadratic modules. *[Hint: if Q is isometric to Q', then show that there is a similarity on the trace zero elements of the duals, thereby giving an isomorphism $\mathcal{O} \xrightarrow{\sim} \mathcal{O}'$].*

13. Let $Q \colon M \to L$ be a quadratic module. Show that the even Clifford algebra $\mathrm{Clf}^0(Q)$ with its map $\iota : M \otimes M \otimes L^\vee \to \mathrm{Clf}^0(Q)$ has the following universal property: if A is an R-algebra and $\iota_A : M \otimes M \otimes L \to A$ is an R-module homomorphism such that

    (i) $\iota_A(x \otimes x \otimes f) = f(Q(x))$ for all $x \in M$ and $f \in L^\vee$, and
    (ii) $\iota_A(x \otimes y \otimes f)\iota_A(y \otimes z \otimes g) = f(Q(y))\iota_A(x \otimes z \otimes g)$ for all $x, y, z \in M$ and $f, g \in L^\vee$,

    then there exists a unique R-algebra homomorphism $\phi : \mathrm{Clf}^0(Q) \to A$ such that the diagram

$$\begin{array}{ccc} M \otimes M \otimes L^\vee & \xrightarrow{\iota} & \mathrm{Clf}(Q) \\ & \searrow^{\iota_A} & \downarrow \phi \\ & & A \end{array}$$

    commutes. Conclude that the pair $(\mathrm{Clf}^0(Q), \iota)$ is unique up to unique isomorphism.

This chapter is licensed under the terms of the Creative Commons Attribution-NonCommercial 4.0 International License (http://creativecommons.org/licenses/by-nc/4.0/), which permits any noncommercial use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license and indicate if changes were made.

The images or other third party material in this chapter are included in the chapter's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the chapter's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.

Chapter 23

Quaternion orders

In the previous chapter, we gave a rather general classification of quaternion orders in terms of ternary quadratic modules. In this chapter, we take a guided tour of the most important animals in the zoo of quaternion orders, identifying those with good local properties. We continue in the next chapter with a second visit to the zoo.

23.1 ▹ Highlights of quaternion orders

We begin in this section by providing some highlights of this tour. Let B be a quaternion algebra over $\mathbb{Q}$ of discriminant $D := \operatorname{disc} B$ and let $\mathcal{O} \subset B$ be an order with reduced discriminant $N := \operatorname{discrd}(\mathcal{O})$. Then $N = DM$ with $M \in \mathbb{Z}_{\geq 1}$, and $\mathcal{O}$ is maximal if and only if $N = D$.

23.1.1 (Maximal orders). The nicest orders are undoubtedly the maximal orders, those not properly contained in another order. An order is maximal if and only if it is locally maximal (Lemma 10.4.3), i.e. p-maximal for all primes p; globally, an order $\mathcal{O}$ is maximal if and only if $N = D$ (i.e., $M = 1$).

We have either $B \simeq \mathrm{M}_2(\mathbb{Q}_p)$ or B is a division algebra over $\mathbb{Q}_p$ (unique up to isomorphism). If B is split, then a maximal order is isomorphic (conjugate) to $\mathrm{M}_2(\mathbb{Z}_p)$, and the corresponding ternary quadratic form is the determinant $xy - z^2$ (see Example 22.3.12). If instead B is division, then the unique maximal order is the valuation ring, with corresponding anisotropic form $x^2 - ey^2 + pz^2$ for $p \neq 2$, where $e \in \mathbb{Z}$ is a quadratic nonresidue modulo p (and for $p = 2$, the associated form is $x^2 + xy + y^2 + 2z^2$).

Maximal orders have modules with good structural properties: all lattices $I \subset B$ with left or right order equal to a maximal order $\mathcal{O}$ are invertible (Theorem 18.1.2).

There is a combinatorial structure, called the **Bruhat–Tits tree**, that classifies maximal orders in $\mathrm{M}_2(\mathbb{Q}_p)$ (as endomorphism rings of lattices, up to scaling): the Bruhat–Tits tree is a $p + 1$-regular tree (see section 23.5).

Examining orders beyond maximal orders is important for the development of the theory: already the Lipschitz order—an order which arises when considering if a

© The Author(s) 2021

J. Voight, *Quaternion Algebras*, Graduate Texts in Mathematics 288,
https://doi.org/10.1007/978-3-030-56694-4_23

positive integer is the sum of four squares—is properly contained inside the Hurwitz order (Chapter 11).

23.1.2 (Hereditary orders). More generally, we say that O is **hereditary** if every left or right fractional O-ideal (i.e., lattice $I \subseteq B$ with left or right order containing O) is invertible. Maximal orders are hereditary, and being hereditary is a local property. A hereditary $\mathbb{Z}_p$-order $O_p \subseteq B_p$ is either maximal or

$$O_p \simeq \begin{pmatrix} \mathbb{Z}_p & \mathbb{Z}_p \\ p\mathbb{Z}_p & \mathbb{Z}_p \end{pmatrix} = \left\{ \begin{pmatrix} a & b \\ pc & d \end{pmatrix} : a,b,c,d \in \mathbb{Z}_p \right\} \subseteq \mathrm{M}_2(\mathbb{Q}_p) \simeq B_p$$

with associated ternary quadratic form $xy - pz^2$. Thus $O \subset B$ is hereditary if and only if $\operatorname{discrd}(O) = DM$ is squarefree, so in particular $\gcd(D, M) = 1$.

Hereditary orders share the nice structural property of maximal orders: all lattices $I \subset B$ with hereditary left or right order are invertible. The different ideal $\operatorname{diff} O_p$ is generated by any element $\mu \in O_p$ such that $\mu^2 \in p\mathbb{Z}_p$.

23.1.3 (Eichler orders). More generally, we can consider orders that are “upper triangular modulo M” with $\gcd(D, M) = 1$ (i.e., avoiding primes that ramify in B). The order

$$\begin{pmatrix} \mathbb{Z}_p & \mathbb{Z}_p \\ p^e\mathbb{Z}_p & \mathbb{Z}_p \end{pmatrix} \subseteq \mathrm{M}_2(\mathbb{Z}_p)$$

is called the **standard Eichler order of level** p^e in $\mathrm{M}_2(\mathbb{Q}_p)$. A $\mathbb{Z}_p$-order $O_p \subseteq \mathrm{M}_2(\mathbb{Q}_p)$ is an **Eichler order** if O_p is isomorphic to a standard Eichler order. The ternary quadratic form associated to an Eichler order of level p^e is $xy - p^e z^2$.

Globally, we say $O \subset B$ is a **Eichler order of level** M if $\operatorname{discrd}(O) = N = DM$ with $\gcd(D, M) = 1$ and O_p is an Eichler order of level p^e for all $p^e \parallel M$. In particular, O_p is maximal at all primes $p \mid D$. Every hereditary order is Eichler, and an Eichler order is hereditary if and only if its level M (or reduced discriminant N) is squarefree. A maximal $\mathbb{Z}_p$-order $O_p \subseteq \mathrm{M}_2(\mathbb{Z}_p)$ is an Eichler order of level $1 = p^0$. Eichler orders play a crucial role in the context of modular forms, as we will see in the final part of this monograph.

This local description of Eichler orders also admits a global description. The standard Eichler order O_p can be written

$$O_p = \begin{pmatrix} \mathbb{Z}_p & \mathbb{Z}_p \\ p^e\mathbb{Z}_p & \mathbb{Z}_p \end{pmatrix} = \mathrm{M}_2(\mathbb{Z}_p) \cap \begin{pmatrix} \mathbb{Z}_p & p^{-e}\mathbb{Z}_p \\ p^e\mathbb{Z}_p & \mathbb{Z}_p \end{pmatrix} = \mathrm{M}_2(\mathbb{Z}_p) \cap \varpi^{-1}\,\mathrm{M}_2(\mathbb{Z}_p)\varpi$$

as the intersection of a (unique) pair of maximal orders, with

$$\varpi := \begin{pmatrix} 0 & 1 \\ p^e & 0 \end{pmatrix} \in N_{\mathrm{GL}_2(\mathbb{Q}_p)}(O_p) \tag{23.1.4}$$

a generator of the group $N_{\mathrm{GL}_2(\mathbb{Q}_p)}(O_p)/\mathbb{Q}_p^\times O_p^\times \simeq \mathbb{Z}/2\mathbb{Z}$, and $\varpi^2 = p^e$. The different $\operatorname{diff} O_p$ is the two-sided ideal generated by ϖ.

From the local-global dictionary, it follows that $O \subset B$ is Eichler if and only if O is the intersection of two (not necessarily distinct) maximal orders.

23.2 Maximal orders

Throughout this chapter, we impose the following notation: let R be a Dedekind domain with field of fractions $F = \operatorname{Frac} R$, let B be a quaternion algebra over F, and let $O \subseteq B$ an R-order.

23.2.1. We make the following convention. When we say "R is local", we mean that R is a complete DVR, and in this setting we let $\mathfrak{p} = \pi R$ be its maximal ideal, and $k = R/\mathfrak{p}$ the residue field. When we want to return to the general context, we will say "R is Dedekind".

Recall that an R-order is *maximal* if it is not properly contained in another order. We begin in this section by summarizing the properties of maximal orders for convenience.

23.2.2. Being maximal is a local property (Lemma 10.4.3), so the following are equivalent:

(i) O is a maximal R-order;
(ii) $O_{(\mathfrak{p})}$ is a maximal $R_{(\mathfrak{p})}$-order for all $\mathfrak{p} \subseteq R$; and
(iii) $O_{\mathfrak{p}}$ is a maximal $R_{\mathfrak{p}}$-order for all $\mathfrak{p} \subseteq R$.

We recall (Theorem 15.5.5) that an order over a global ring R is maximal if and only if $\operatorname{discrd}(O) = \operatorname{disc}_R(B)$. Local maximal R-orders have the nice local description.

23.2.3. Suppose R is local and that $B \simeq \mathrm{M}_2(F)$ is split. Then by Corollary 10.5.5, every maximal R-order in $\mathrm{M}_2(F)$ is conjugate to $\mathrm{M}_2(R)$ by an element of $\mathrm{GL}_2(R)$, i.e. $O \simeq \mathrm{M}_2(R)$. We have $\operatorname{discrd}(O) = R$. All two-sided ideals of O are powers of $\operatorname{rad}(O) = \mathfrak{p}O$, and

$$O/\operatorname{rad}(O) \simeq \mathrm{M}_2(k).$$

The associated ternary quadratic form is similar to $Q(x, y, z) = xy - z^2$, by Example 22.3.12 and the classification theorem (Main Theorem 22.1.1). Finally,

$$N_{B^\times}(O) = N_{\mathrm{GL}_2(F)}(\mathrm{M}_2(R)) = F^\times O^\times. \tag{23.2.4}$$

23.2.5. Suppose R is local but now that B is a division algebra. Then the valuation ring $O \subset B$ is the unique maximal R-order by Proposition 13.3.4.

Suppose further that the residue field k is finite (equivalently, that F is a local field). Then Theorem 13.3.11 applies, and we have

$$O \simeq S \oplus Sj \subseteq \left(\frac{K, \pi}{F}\right)$$

where $K \supseteq F$ is the unique quadratic unramified extension of F and S its valuation ring. We computed in 15.2.12 that $\operatorname{discrd}(O) = \mathfrak{p}$. All two-sided ideals of O are powers of the unique maximal ideal $\operatorname{rad}(O) = P = OjO$, and $\ell := O/\operatorname{rad}(O)$ is a quadratic field extension of k. By Exercise 13.7, we have $P = [O, O]$ equal to the commutator. We also have $P = \operatorname{diff} O$; this can be computed directly, or it follows from the condition that $\operatorname{nrd}(\operatorname{diff} O) = \operatorname{discrd} O = \mathfrak{p}$.

Write $S = R[i]$ with $i^2 = ui - b$, and $u, b \in R$. Then $1, i, j, k$ where $k = -ij$ is an R-basis for O. We have

$$k^2 = i(ji)j = i(\bar{i}j)j = \operatorname{nrd}(i)\pi,$$

so $\operatorname{trd}(k) = 0$, and $\bar{k} = -k = ij$. This gives multiplication table

$$\begin{aligned} i^2 &= ui - b & jk &= -\pi\bar{i} \\ j^2 &= \pi & ki &= b\bar{j} \\ k^2 &= b\pi & ij &= \bar{k} \end{aligned} \tag{23.2.6}$$

realizing the basis as a good basis in the sense of 22.4.7; the associated ternary quadratic form is

$$\operatorname{nrd}^\sharp(O)(x, y, z) = -\pi x^2 + by^2 + uyz + z^2 = -\pi x^2 + \operatorname{Nm}_{K|F}(z + yi). \tag{23.2.7}$$

Finally, since the valuation ring is the unique maximal order and conjugation respects integrality, we have

$$N_{B^\times}(O) = B^\times. \tag{23.2.8}$$

Finally, lattices over maximal orders are necessarily invertibility, as follows.

23.2.9. All lattices $I \subset B$ with left or right order equal to a maximal order O are invertible, by Theorem 18.1.2, proven in Proposition 18.3.2. (We also gave a different proof of this fact in Proposition 16.6.15(b).)

The classification of two-sided ideals and their classes follows from that of hereditary orders: see 23.3.19.

23.3 Hereditary orders

Hereditary orders were investigated in section 21.4 in general; here, we provide a quick development specific to quaternion algebras. As mentioned before, a good general reference for (maximal and) hereditary orders is Reiner [Rei2003, Chapters 3–6, 9].

We recall that O is *hereditary* if every left (or right) ideal $I \subseteq O$ is projective as a left (or right) O-module. Being hereditary is a local property 21.4.4 because projectivity is.

23.3.1. Suppose R is local. By Main Theorem 21.1.4 and Corollary 21.1.5, the following are equivalent:

(i) O is hereditary;

(ii) $\operatorname{rad} O$ is projective as a left (or right) O-module;

(iii) $O_{\mathsf{L}}(\operatorname{rad} O) = O_{\mathsf{R}}(\operatorname{rad} O) = O$;

(iv) $\operatorname{rad} O$ is invertible as a (sated) two-sided O-ideal; and

(v) either O is maximal or

$$O \simeq \begin{pmatrix} R & R \\ \mathfrak{p} & R \end{pmatrix} \subseteq \mathrm{M}_2(F) \simeq B.$$

We now spend some time investigating 'the' local hereditary order that is not maximal. So until further notice, suppose R is local and let

$$O_0(\mathfrak{p}) := \begin{pmatrix} R & R \\ \mathfrak{p} & R \end{pmatrix} \subseteq \mathrm{M}_2(R).$$

To avoid clutter, we will just write $O = O_0(\mathfrak{p})$. We have

$$\operatorname{discrd}(O) = [\mathrm{M}_2(R) : O]_R \operatorname{disc}(\mathrm{M}_2(R)) = \mathfrak{p}. \tag{23.3.2}$$

23.3.3. A multiplication table for O is obtained from the one for $\mathrm{M}_2(R)$ in Example 22.3.12, scaling j by π in (22.3.12), which gives the same multiplication laws as (22.3.13) except now $ij = -\pi\overline{k}$ and c is scaled by π. Therefore, the similarity class of ternary quadratic forms associated to O is represented by

$$Q(x, y, z) = xy - \pi z^2. \tag{23.3.4}$$

23.3.5. Let $J := \operatorname{rad}(O)$. Then

$$J = \begin{pmatrix} \mathfrak{p} & R \\ \mathfrak{p} & \mathfrak{p} \end{pmatrix} \tag{23.3.6}$$

by (21.3.6), and we find

$$O/J \simeq \begin{pmatrix} k & 0 \\ 0 & k \end{pmatrix} \simeq k \times k \tag{23.3.7}$$

as k-algebras. Now let

$$\varpi = \begin{pmatrix} 0 & 1 \\ \pi & 0 \end{pmatrix}. \tag{23.3.8}$$

Then a direct calculation yields

$$J = O\varpi = \varpi O \tag{23.3.9}$$

in agreement with 23.3.1(ii)–23.3.1(iii), and J is an invertible O-ideal. Since $\varpi^2 = \pi$, we have

$$J^2 = \mathfrak{p}O. \tag{23.3.10}$$

In particular, $J^{-1} = \pi^{-1}J$, and the powers of J give a filtration

$$O \supsetneq J \supsetneq \mathfrak{p}O \supsetneq J^3 \supsetneq \dots. \tag{23.3.11}$$

23.3.12. We compute that

$$\varpi\, \mathrm{M}_2(R)\varpi^{-1} = \begin{pmatrix} R & \mathfrak{p}^{-1}R \\ \mathfrak{p} & R \end{pmatrix}$$

and hence

$$O = \mathrm{M}_2(R) \cap \varpi\, \mathrm{M}_2(R)\varpi^{-1}$$

is the intersection of two maximal orders.

Lemma 23.3.13. *The group* $\mathrm{Idl}(O) = \mathrm{PIdl}(O)$ *is generated by* $J = \operatorname{rad} O$*, with* $J^2 = \mathfrak{p}O$*.*

Proof. We have $\mathrm{Idl}(O) = \mathrm{PIdl}(O)$ since R is local: invertible is equivalent to principal (Main Theorem 16.6.1).

Let $I \subseteq O$ be an invertible two-sided O-ideal. Then by (23.3.11), we can replace I by a power of J^{-1} and suppose that $O \subsetneq I \subseteq J$. Invertible means locally principal, so $I = O\alpha = \alpha O$ and

$$[O : I]_R = \mathrm{nrd}(I)^2 = \mathrm{nrd}(\alpha)^2 R \mid [O : J]_R = \mathfrak{p}^2.$$

Thus $[O : I]_R = \mathfrak{p}^2$, so $[I : J]_R = R$ and $I = J$. (This also follows directly from Proposition 16.4.3.)

Here is a second computational proof. The image $I/J \subseteq O/J \simeq k \times k$ (23.3.7) is a two-sided ideal, therefore

$$I = \begin{pmatrix} R & \mathfrak{p} \\ \mathfrak{p} & \mathfrak{p} \end{pmatrix} \quad \text{or} \quad I = \begin{pmatrix} \mathfrak{p} & \mathfrak{p} \\ \mathfrak{p} & R \end{pmatrix}.$$

But

$$\begin{pmatrix} R & \mathfrak{p} \\ \mathfrak{p} & \mathfrak{p} \end{pmatrix}\begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix} = \begin{pmatrix} 0 & R \\ 0 & \mathfrak{p} \end{pmatrix} \not\subseteq \begin{pmatrix} R & \mathfrak{p} \\ \mathfrak{p} & \mathfrak{p} \end{pmatrix};$$

we get a contradiction with the other possibility by multiplying instead on the left. A third "pure matrix multiplication" proof is also requested in Exercise 23.1.

The second statement was already proven in (23.3.10). □

Corollary 23.3.14. *We have* $N_{B^\times}(O)/(F^\times O^\times) \simeq \mathbb{Z}/2\mathbb{Z}$ *generated by* ϖ*, and*

$$\mathrm{nrd}(N_{B^\times}(O)) = \begin{cases} F^{\times 2} R_{\mathfrak{p}}^\times, & \text{if } e \text{ is even;} \\ F^\times, & \text{if } e \text{ is odd.} \end{cases}$$

Proof. By (18.5.4), we have an isomorphism

$$N_{B^\times}(O)/(F^\times O^\times) \simeq \mathrm{PIdl}(O)/\,\mathrm{PIdl}(R);$$

by Lemma 23.3.13, the latter is generated by $J = \varpi O$ with $J^2 = \pi O$. The computation of reduced norms is immediate. □

23.3.15. Lemma 23.3.13 also implies the description

$$J = [O, O] \tag{23.3.16}$$

as the commutator. Since O/J is commutative, we know $[O, O] \subseteq J$; but $[O, O] \subsetneq J^2 = \mathfrak{p}O$ since $O/\mathfrak{p}O$ is noncommutative. We also have $J = \operatorname{diff} O$ for the same reason, since $\mathrm{nrd}(\operatorname{diff} O) = \operatorname{discrd} O = \mathfrak{p}$. (A matrix proof of these facts are requested in Exercise 23.8.)

23.3.17. We now classify the left O-lattices, up to isomorphism. Each such O-lattice is projective, since O is hereditary.

By the Krull-Schmidt theorem (Theorem 20.6.2), every O-lattice can be written as the direct sum of indecomposables, so it is enough to classify the indecomposables; and we did just this in 21.5.3. Explicitly, we have $V = \begin{pmatrix} F \\ F \end{pmatrix} \simeq F^2$ the simple $B = \mathrm{M}_2(F)$-module, and we take the O-lattice $M = \begin{pmatrix} R \\ R \end{pmatrix} \subset V$. We have $JM = \begin{pmatrix} R \\ \mathfrak{p} \end{pmatrix}$ and $J^2M = \begin{pmatrix} \mathfrak{p} \\ \mathfrak{p} \end{pmatrix} = \pi M$, and M, JM give a complete set of indecomposable left O-modules. As expected, $O = JM \oplus M$ is a decomposition of O into projective indecomposable left O-modules.

The preceding local results combine to determine global structure. Now let R be a global ring.

Lemma 23.3.18. *O is hereditary if and only if* $\operatorname{discrd}(O)$ *is squarefree.*

Proof. We argue locally; and then we use the characterization (iv), the computation of the reduced discriminant (23.3.2), and the same argument as in Theorem 15.5.5 to finish. $\square$

23.3.19. Let O be a hereditary (possibly maximal) R-order. By Theorem 21.4.9, we know that the group $\operatorname{Idl}(O)$ is an abelian group generated by the prime (equivalently, maximal) invertible two-sided ideals. We claim that the map

$$\begin{gathered}\{\text{Prime two-sided invertible } O-\text{ ideals}\} \leftrightarrow \{\text{Prime ideals of } R\} \\ P \mapsto P \cap R\end{gathered} \tag{23.3.20}$$

is a bijection, generalizing Theorem 18.3.6. If $\mathfrak{p} \nmid \mathfrak{N}$ then we have $P = \mathfrak{p}O$; and if $\mathfrak{p} \mid \mathfrak{D}$ then we have a prime two-sided ideal $P = O \cap \operatorname{rad}(O_{\mathfrak{p}})$ with $P^2 = \mathfrak{p}O$. Otherwise, $\mathfrak{p} \mid \mathfrak{N}$ but $\mathfrak{p} \nmid \mathfrak{D}$, so $O_{\mathfrak{p}}$ is hereditary but not maximal; from the local description in Lemma 23.3.13, we get a prime ideal $P = O \cap \operatorname{rad}(O_{\mathfrak{p}})$ with $P^2 = \mathfrak{p}O$ as in the ramified case. This proves (23.3.20), and that the sequence

$$0 \to \operatorname{Idl}(R) \to \operatorname{Idl}(O) \to \prod_{\mathfrak{p} \mid \mathfrak{N}} \mathbb{Z}/2\mathbb{Z} \to 0 \tag{23.3.21}$$

is exact.

Taking the quotient by $\operatorname{PIdl}(R)$, we obtain the exact sequence

$$0 \to \operatorname{Cl} R \to \operatorname{Pic}_R(O) \to \prod_{\mathfrak{p} \mid \mathfrak{N}} \mathbb{Z}/2\mathbb{Z} \to 0. \tag{23.3.22}$$

In particular, if $\mathfrak{D} = (1)$, then $\operatorname{Pic}_R(O) \simeq \operatorname{Cl} R$. Finally, the group of two-sided ideals modulo principal two-sided ideals is related to the Picard group by the exact sequence (18.5.5):

$$\begin{gathered}0 \to N_{B^\times}(O)/(F^\times O^\times) \to \operatorname{Pic}_R(O) \to \operatorname{Idl}(O)/\operatorname{PIdl}(O) \to 0 \\ \alpha F^\times O^\times \mapsto [\alpha O] = [O\alpha]\end{gathered}$$

(This exact sequence is sensitive to O even within its genus: see Remark 18.5.9.)

23.4 Eichler orders

We now consider a more general class of orders inspired by the hereditary orders.

Definition 23.4.1. An **Eichler order** $\mathcal{O} \subseteq B$ is the intersection of two (not necessarily distinct) maximal orders.

23.4.2. By the local-global dictionary for lattices (and orders), the property of being an Eichler order is local. Moreover, from 23.3.12, it follows that a hereditary order is Eichler.

Proposition 23.4.3. *Suppose R is local and $\mathcal{O} \subseteq B = \mathrm{M}_2(F)$. Then the following are equivalent:*

(i) *$\mathcal{O}$ is Eichler;*

(ii) *$\mathcal{O} \simeq \begin{pmatrix} R & R \\ \mathfrak{p}^e & R \end{pmatrix}$;*

(iii) *$\mathcal{O}$ contains an R-subalgebra that is $B^\times$-conjugate to $\begin{pmatrix} R & 0 \\ 0 & R \end{pmatrix}$; and*

(iv) *$\mathcal{O}$ is the intersection of a uniquely determined pair of maximal orders (not necessarily distinct).*

Proof. We follow Hijikata [Hij74, 2.2(i)]. Apologies in advance for all of the explicit matrix multiplication!

We prove (i) $\Rightarrow$ (ii) $\Leftrightarrow$ (iii) and then (ii) $\Rightarrow$ (iv) $\Rightarrow$ (i). The implications (ii) $\Rightarrow$ (iii) and (iv) $\Rightarrow$ (i) are immediate.

So first (i) $\Rightarrow$ (ii). Suppose $\mathcal{O} = \mathcal{O}_1 \cap \mathcal{O}_2$. All maximal orders in B are $B^\times$-conjugate to $\mathrm{M}_2(R)$, so there exist $\alpha_1, \alpha_2 \in B^\times$ such that $\mathcal{O}_i = \alpha_i^{-1}\,\mathrm{M}_2(R)\alpha_i$ for $i = 1, 2$. Conjugating by α_1, we may suppose $\alpha_1 = 1$ and we write $\alpha = \alpha_2^{-1}$ for convenience, so $\mathcal{O} \simeq \mathrm{M}_2(R) \cap \alpha\,\mathrm{M}_2(R)\alpha^{-1}$. Scaling by π, we may suppose $\alpha \in \mathrm{M}_2(R) \smallsetminus \pi\,\mathrm{M}_2(R)$. By row and column operations (Smith normal form, proven as part of the structure theorem for finitely generated modules over a PID), there exist $\beta, \gamma \in \mathrm{GL}_2(R)$ such that

$$\beta\alpha\gamma = \begin{pmatrix} 1 & 0 \\ 0 & \pi^e \end{pmatrix}$$

is in standard invariant form with $e \geq 0$. Then

$$\mathcal{O} \simeq \beta\mathcal{O}\beta^{-1} = \mathrm{M}_2(R) \cap \beta\alpha\,\mathrm{M}_2(R)\alpha^{-1}\beta^{-1} \tag{23.4.4}$$

since $\beta \in \mathrm{GL}_2(R)$, and

$$\beta\alpha\,\mathrm{M}_2(R)\alpha^{-1}\beta^{-1} = \begin{pmatrix} 1 & 0 \\ 0 & \pi^e \end{pmatrix}\begin{pmatrix} R & R \\ R & R \end{pmatrix}\begin{pmatrix} 1 & 0 \\ 0 & \pi^{-e} \end{pmatrix} = \begin{pmatrix} R & \mathfrak{p}^{-e} \\ \mathfrak{p}^e & R \end{pmatrix} \tag{23.4.5}$$

so

$$\mathcal{O} \simeq \begin{pmatrix} R & R \\ R & R \end{pmatrix} \cap \begin{pmatrix} R & \mathfrak{p}^{-e} \\ \mathfrak{p}^e & R \end{pmatrix} = \begin{pmatrix} R & R \\ \mathfrak{p}^e & R \end{pmatrix}. \tag{23.4.6}$$

To show (iii) ⇒ (ii), we may suppose $O \supseteq \begin{pmatrix} R & 0 \\ 0 & R \end{pmatrix} = Re_{11} + Re_{22}$ with $e_{11} = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}, e_{22} = \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix}$. Then $Oe_{11} \subseteq \begin{pmatrix} F & 0 \\ F & 0 \end{pmatrix}$. Let π_{ij} be the projection onto the ij-coordinate. Then

$$\pi_{11}(Oe_{11}) = \operatorname{trd}(Oe_{11})e_{11} \subseteq R = \begin{pmatrix} R & 0 \\ 0 & 0 \end{pmatrix}$$

and so equality holds. Therefore

$$O \supseteq \pi_{21}(O) = \begin{pmatrix} 0 & 0 \\ \mathfrak{p}^a & 0 \end{pmatrix}$$

for some $a \in \mathbb{Z}$. Arguing again with the other matrix unit e_{22} we conclude that

$$O = \begin{pmatrix} R & \mathfrak{p}^b \\ \mathfrak{p}^a & R \end{pmatrix} \tag{23.4.7}$$

with $a, b \in \mathbb{Z}$. Multiplying

$$\begin{pmatrix} R & \mathfrak{p}^b \\ \mathfrak{p}^a & R \end{pmatrix} \begin{pmatrix} R & \mathfrak{p}^b \\ \mathfrak{p}^a & R \end{pmatrix} = \begin{pmatrix} R + \mathfrak{p}^{a+b} & \mathfrak{p}^b \\ \mathfrak{p}^a & R + \mathfrak{p}^{a+b} \end{pmatrix}$$

we conclude that $e = a + b \geq 0$. Such an order is maximal if and only if $a + b = 0$: if $a \geq 0$, then $\begin{pmatrix} R & \mathfrak{p}^b \\ \mathfrak{p}^a & R \end{pmatrix} \subseteq \begin{pmatrix} R & \mathfrak{p}^{-a} \\ \mathfrak{p}^a & R \end{pmatrix}$ and similarly if $a \leq 0$. The element $\alpha = \begin{pmatrix} 0 & 1 \\ \pi^a & 0 \end{pmatrix}$ has

$$\alpha^{-1} O \alpha = \begin{pmatrix} 0 & 1 \\ \pi^a & 0 \end{pmatrix} \begin{pmatrix} R & \mathfrak{p}^b \\ \mathfrak{p}^a & R \end{pmatrix} \begin{pmatrix} 0 & \pi^{-a} \\ 1 & 0 \end{pmatrix} = \begin{pmatrix} R & R \\ \mathfrak{p}^e & R \end{pmatrix} \tag{23.4.8}$$

(and normalizes the given subalgebra) so the result is proven.

To conclude, we show (ii) ⇒ (iv). Let $O' \supseteq O = \begin{pmatrix} R & R \\ \mathfrak{p}^e & R \end{pmatrix}$ be a maximal R-order. Since $\begin{pmatrix} R & 0 \\ 0 & R \end{pmatrix} \subseteq O'$, the argument of the previous paragraph applies, and $O' = \begin{pmatrix} R & \mathfrak{p}^{-c} \\ \mathfrak{p}^c & R \end{pmatrix}$ with $c \in \mathbb{Z}$ satisfying $0 \leq c \leq e$. The intersection of another such maximal orders with the parameter d is the order $\begin{pmatrix} R & \mathfrak{p}^a \\ \mathfrak{p}^b & R \end{pmatrix}$ where $a = \max(c, d)$ and $b = -\min(c, d)$ so is equal to $\begin{pmatrix} R & R \\ \mathfrak{p}^e & R \end{pmatrix}$ if and only if $e = a = \max(c, d)$ and $0 = b = \min(c, d)$, which uniquely determine c, d up to swapping. □

Remark 23.4.9. There is a further important equivalent characterization of Eichler orders as being maximal or *residually split*: see Lemma 24.3.6.

Corollary 23.4.10. *Every superorder of an Eichler order is Eichler.*

Proof. The corollary is local, so we may apply Proposition 23.4.3(iii) to every superorder. ☐

Definition 23.4.11. Suppose R is local. The **standard Eichler order of level** $\mathfrak{p}^e$ in $\mathrm{M}_2(F)$ is the order

$$O_0(\mathfrak{p}^e) := \begin{pmatrix} R & R \\ \mathfrak{p}^e & R \end{pmatrix}.$$

By Proposition 23.4.3, if R is local then an order $O \subseteq \mathrm{M}_2(F)$ is Eichler if and only if O is conjugate to a standard Eichler order.

Suppose until further notice that R is local, and let $O = O_0(\mathfrak{p}^e)$ be the standard Eichler order of level $\mathfrak{p}^e$ with $e \geq 0$.

23.4.12. First two basic facts about the Eichler order O of level $\mathfrak{p}^e$: We have

$$\operatorname{discrd}(O) = [\mathrm{M}_2(R) : O]_R = \mathfrak{p}^e$$

and its associated ternary quadratic form $Q(x, y, z) = xy - \pi^e z^2$ as in 23.3.3.

23.4.13. Let

$$\varpi = \begin{pmatrix} 0 & 1 \\ \pi^e & 0 \end{pmatrix} \in O.$$

Then

$$O = \mathrm{M}_2(R) \cap \varpi^{-1} \mathrm{M}_2(R) \varpi$$

as in (23.4.6); by Proposition 23.4.3 these two orders are the uniquely determined pair of maximal orders containing O. We have $\varpi^2 = \pi^e$, and so $\varpi \in N_{B^\times}(O)$. It follows (and can be checked directly) that $I = O\varpi = \varpi O$ is a two-sided O-ideal. If $e = 0$, we have $I = O$.

Proposition 23.4.14. *Suppose that $e \geq 1$. Then we have $N_{B^\times}(O)/(F^\times O^\times) = \langle \varpi \rangle \simeq \mathbb{Z}/2\mathbb{Z}$. Moreover, the group $\operatorname{Idl}(O) = \operatorname{PIdl}(O)$ is abelian, generated by I and $\mathfrak{p}O$ with the single relation $I^2 = \mathfrak{p}^e O$.*

Proof. Let $\alpha \in N_{B^\times}(O)$. Then by uniqueness of the intersection in 23.4.13, conjugation by α permutes these two orders, so we have a homomorphism $N_{B^\times}(O)$ to a cyclic group of order 2. This homomorphism is surjective, since ϖ transposes the orders. If α is in the kernel, then $\alpha \in N_{B^\times}(\mathrm{M}_2(R)) = F^\times \mathrm{GL}_2(R)$ and unconjugating the second factor we similarly get $\varpi\alpha\varpi^{-1} \in F^\times \mathrm{GL}_2(R)$, so

$$\alpha \in F^\times(\mathrm{GL}_2(R) \cap \varpi^{-1} \mathrm{GL}_2(R)\varpi) = F^\times O^\times.$$

Again since R is local, we have $\operatorname{Idl}(O) = \operatorname{PIdl}(O)$, and by (18.5.4), we have an isomorphism

$$N_{B^\times}(O)/(F^\times O^\times) \simeq \operatorname{PIdl}(O)/\operatorname{PIdl}(R)$$

so $\operatorname{Idl}(O)$ is generated by I and the generator $\mathfrak{p}$ for $\operatorname{PIdl}(R)$. ☐

23.4.15. It is helpful to consider the Jacobson radical of an Eichler order, to compare to the hereditary case.

$$J = \begin{pmatrix} \mathfrak{p} & R \\ \mathfrak{p}^e & \mathfrak{p} \end{pmatrix}.$$

We claim that $J = \operatorname{rad} O$. We verify directly that $J \subseteq O$ is a two-sided ideal and

$$J^2 = \begin{pmatrix} \mathfrak{p}^f & \mathfrak{p} \\ \mathfrak{p}^{e+1} & \mathfrak{p}^f \end{pmatrix} \subseteq \mathfrak{p}O, \tag{23.4.16}$$

where $f = \min(e, 2)$, so by Corollary 20.5.5, $J \subseteq \operatorname{rad} O$; on the other hand, the quotient

$$O/J \simeq k \times k \tag{23.4.17}$$

is semisimple, so $\operatorname{rad} O \subseteq J$ by Corollary 20.4.11(a).

However, the radical is *not* an invertible (sated) two-sided O-ideal unless O is hereditary ($e = 1$), by 23.3.1. Indeed, we verify that

$$O_{\mathsf{L}}(J) = \begin{pmatrix} R & \mathfrak{p}^{-1} \\ \mathfrak{p}^{e-1} & R \end{pmatrix} = O_{\mathsf{R}}(J) \tag{23.4.18}$$

(Exercise 23.6); this recovers O if and only if $e = 1$, and if $e \geq 2$ then it is an Eichler order of level $\mathfrak{p}^{e-2}$ (conjugating as in (23.4.8)). By (23.4.16), if $e \geq 2$ then $J^2 = \mathfrak{p}J$, and so we certainly could not have J invertible!

We now repackage these local efforts into a global characterization.

23.4.19. Suppose that R is a global ring. Let $\operatorname{disc}_R B = \mathfrak{D}$ and let O be an Eichler order with $\operatorname{discrd} O = \mathfrak{N}$. If $\mathfrak{p} \mid \mathfrak{D}$, then $B_{\mathfrak{p}}$ has a unique maximal order, so (as an 'intersection') $O_{\mathfrak{p}}$ is necessarily the maximal order. If $\mathfrak{p} \nmid \mathfrak{D}$, and $\operatorname{ord}_{\mathfrak{p}} \mathfrak{N} = e \geq 0$, then $O_{\mathfrak{p}}$ is isomorphic to the standard Eichler order of level $\mathfrak{p}^e$.

We have $\mathfrak{N} = \mathfrak{D}\mathfrak{M}$ with $\mathfrak{M} \subseteq R$ and we just showed that $\mathfrak{M}$ is coprime to $\mathfrak{D}$. We call $\mathfrak{M}$ the **level** of the Eichler order O. The pair $\mathfrak{D}, \mathfrak{M}$ (or $\mathfrak{D}, \mathfrak{N}$) determines a unique genus of Eichler R-orders, i.e., this data uniquely determines the isomorphism class of $O_{\mathfrak{p}}$ for each $\mathfrak{p}$.

Putting together Proposition 23.4.14 together with 23.3.19 for the remaining primes where the order is maximal, we have an exact sequence

$$0 \to \operatorname{Idl}(R) \to \operatorname{Idl}(O) \to \prod_{\mathfrak{p} \mid \mathfrak{N}} \mathbb{Z}/2\mathbb{Z} \to 0 \tag{23.4.20}$$

and we may take the quotient by $\operatorname{PIdl} R$ to get

$$0 \to \operatorname{Cl} R \to \operatorname{Pic}_R(O) \to \prod_{\mathfrak{p} \mid \mathfrak{N}} \mathbb{Z}/2\mathbb{Z} \to 0 \tag{23.4.21}$$

Remark 23.4.22. Eichler [Eic56a] developed his orders in detail for prime level, and employed them in the study of modular correspondences [Eic56c] and the trace formula [Eic73] in the case of squarefree level. (As mentioned in Remark 24.1.5, Eichler's

earlier work [Eic36, §6] over $\mathbb{Q}$ included some more general investigations, but his later work seemed mostly confined to the hereditary orders.) Hijikata [Hij74, §2.2] later studied these orders in the attempt to generalize Eichler's result beyond the squarefree case; calling the orders *split* (like our name, *residually split*). Pizer [Piz73, p. 77] may be the first who explicitly called them *Eichler* orders.

Remark 23.4.23. An R-order $O \subseteq \mathrm{M}_n(F)$ that contains a $\mathrm{GL}_n(F)$-conjugate of the diagonal matrices $\mathrm{diag}(R, \dots, R)$ (equivalently, containing n orthogonal idempotents) is said to be **tiled**. By 23.4.3, an order $O \subseteq \mathrm{M}_2(F)$ is tiled if and only if it is Eichler. Tiled orders also go by other names (including *graduated orders*) and they arise naturally in many contexts, including representation theory [Pl83] and modular forms.

23.5 Bruhat–Tits tree

In the previous section, we examined Eichler orders as the intersection of two maximal orders. There is a beautiful and useful combinatorial construction—a tree—which keeps track of the containments among maximal orders in aggregate, as follows. For further reading, see Serre [Ser2003, §II.1].

Let F be a nonarchimedean local field with valuation ring R, maximal ideal $\mathfrak{p} = \pi R$, and residue field $k := R/\mathfrak{p}$, and let $q := \#k$. Let $B := \mathrm{M}_2(F)$, and let $V := F^2$ as column vectors, so that $B := \mathrm{End}_F(V)$ acts on the left.

Recalling again section 10.5, every maximal order $O \subset B$ has $O = \mathrm{End}_R(M)$ where $M \subset V$ is an R-lattice. So to understand maximal orders, it is equivalent to understand lattices (and their containments) and the positioning of one lattice inside another.

Lemma 23.5.1. *Let $L, M \subset V$ be R-lattices. Then there exists an R-basis x_1, x_2 of L such that $\pi^{f_1} x_1, \pi^{f_2} x_2$ is an R-basis of M with $f_1, f_2 \in \mathbb{Z}$ and $f_1 \leq f_2$.*

Proof. Exercise 23.7. □

Lemma 23.5.2. *We have $\mathrm{End}_R(L) = \mathrm{End}_R(M)$ if and only if there exists $a \in F^\times$ such that $M = aL$.*

Proof. If $M = aL$ for $a \in F^\times$, then $\mathrm{End}_R(L) = \mathrm{End}_R(M)$. Conversely, suppose that $\mathrm{End}_R(L) = \mathrm{End}_R(M)$. Replacing M by aM with $a \in F^\times$, we may suppose without loss of generality that $L \subseteq M$. By Lemma 23.5.1, we may identify $L = R^2$ with the standard basis and $M = \pi^{f_1} e_1 \oplus \pi^{f_2} e_2$ with $f_1, f_2 \in \mathbb{Z}_{\geq 0}$; rescaling again, and interchanging the basis elements if necessary, we may suppose $e_1 = 0$. Then $\mathrm{End}_R(M) = \mathrm{End}_R(L) \simeq \mathrm{M}_2(R)$ implies $f_2 = 0$ and $L = M$. □

With this lemma in mind, we make the following definition (recalling this definition made earlier in the context of algebras).

Definition 23.5.3. Two R-lattices $L, L' \subset V$ are **homothetic** if there exists $a \in F^\times$ such that $L' = aL$.

The relation of homothety is an equivalence relation on the set of R-lattices in V, and we write $[L]$ for the homothety class of L.

23.5.4. Let $L \subset V$ be an R-lattice. In a homothety class of lattices, there is a unique lattice $L' \subseteq L$ in this homothety class satisfying any of the following equivalent conditions:

(i) $L' \subseteq L$ is maximal;
(ii) $L' \not\subseteq \pi L$; and
(iii) L/L' is cyclic as an R-module (has one generator).

These equivalences follow from Lemma 23.5.1: they are equivalent to $f_1 = 0$, and correspond to a maximal scaling of L' by a power of π within L. For such an L', we have $L/L' \simeq R/\pi^f R$ for a unique $f \geq 0$.

Definition 23.5.5. Let $\mathcal{T}$ be the graph whose vertices are homothety classes of R-lattices in V and where an undirected edge joins two vertices (exactly) when there exist representative lattices L, L' for these vertices such that

$$\pi L \subsetneq L' \subsetneq L. \tag{23.5.6}$$

Equivalently, by Lemma 23.5.2, the vertices of $\mathcal{T}$ are in bijection with maximal orders in $B = \mathrm{M}_2(F)$ by $[L] \mapsto \mathrm{End}_R(L)$ for every choice of $L \in [L]$.

23.5.7. The adjacency relation (23.5.6) implies $L' \subsetneq \pi L \subsetneq \pi L'$, so it is sensible to have undirected edges.

A class $[L']$ has an edge to L if and only if the representative L' in 23.5.4 has $f = 1$.

Proposition 23.5.8. *The graph $\mathcal{T}$ is a connected tree such that each vertex has degree $q + 1$.*

Proof. We have $L/\pi L \simeq k^2$, and so the lattices L' satisfying (23.5.6) are in bijection with k-subspaces of dimension 1 in $L/\pi L$; such a subspace is given by a choice of generator up to scaling, so there are exactly $(q^2 - 1)/(q - 1) = q + 1$ such, and each vertex has $q + 1$ adjacent vertices. The graph is connected: given two vertices, we may choose representative lattices L, L' such that $L' \subseteq L$ as in 23.5.4. The quotient L/L' is cyclic, so by induction the lattices $L_i = \pi^i L + L'$ for $i = 0, \dots, f$ have L_i adjacent to L_{i+1}, and $L_0 = L$ and $L_f = L'$, giving a path from $[L]$ to $[L']$.

The following argument comes from Dasgupta–Teitelbaum [DT2008, Proposition 1.3.2]. Suppose there is a nontrivial cycle in $\mathcal{T}$

$$\pi^v L = L_s \subsetneq L_{s-1} \subsetneq \cdots \subsetneq L_1 \subsetneq L_0 = L \tag{23.5.9}$$

so that $v \geq 1$. We may suppose this cycle is minimal, meaning that no intermediate lattices are equivalent. The quotient $L/L_s = L/\pi^v L \simeq (R/\mathfrak{p}^v)^2$ is not cyclic; let i be the largest index such that L/L_i is cyclic but L/L_{i+1} is not. Thus $L/L_{i+1} \simeq R/\mathfrak{p}^i \oplus R/\mathfrak{p}$, and so π^i annihilates L/L_{i+1} and $\pi^i L \subseteq L_{i+1}$. Since L/L_i is cyclic, just as in the previous paragraph, we conclude $L_{i-1} = \pi^{i-1} L + L_i$. Putting these together, we find

that $\pi L_{i-1} = \pi^i L + \pi L_i \subseteq L_{i+1}$, the latter by definition of the adjacency between L_i and L_{i+1}. By adjacency, $L_i \subsetneq \pi L_{i_1} \subseteq L_{i+1}$. And again by adjacency, L_{i+1} is maximal inside L_i, so $\pi L_{i-1} = L_{i+1}$. This contradicts the minimality of the cycle; we conclude that $\mathcal{T}$ has no cycles. □

We call $\mathcal{T}$ the **Bruhat–Tits tree** for $\mathrm{GL}_2(F)$. The Bruhat–Tits tree for $F = \mathbb{Q}_2$ is sketched in Figure 23.5.10. We write $\mathrm{Ver}(\mathcal{T})$ and $\mathrm{Edg}(\mathcal{T})$ for the set of vertices and edges of $\mathcal{T}$.

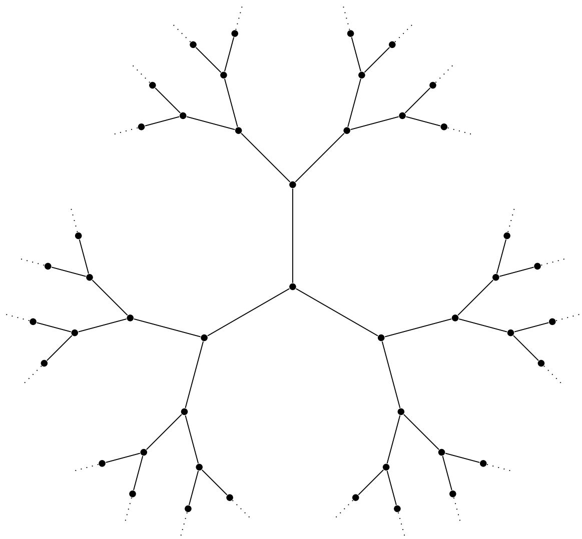


Figure 23.5.10: Bruhat–Tits tree for $\mathrm{GL}_2(\mathbb{Q}_2)$

23.5.11. We define a transitive action of $\mathrm{GL}_2(F)$ on $\mathcal{T}$ as follows.

Let $L \subseteq V$ be a lattice. Choose a R-basis for L and put the vectors in the *columns* of a matrix $\beta \in \mathrm{M}_2(F)$. Since these columns span V over F, we have $\beta \in \mathrm{GL}_2(F)$, and the matrix β is well-defined up to a change of basis over R; therefore the coset $\beta\,\mathrm{GL}_2(R) \in \mathrm{GL}_2(F)/\mathrm{GL}_2(R)$ is well-defined. (Check that the action of change of basis on columns is given by matrix multiplication the right.) Therefore a homothety class $[L]$ gives a well-defined element of $\mathrm{GL}_2(F)/(F^\times \mathrm{GL}_2(R))$. Conversely, given such a class we can consider the R-lattice spanned by its columns, and its homothety class is well-defined. We have shown there is a bijection

$$\mathrm{Ver}(\mathcal{T}) \leftrightarrow \mathrm{GL}_2(F)/(F^\times \mathrm{GL}_2(R)). \tag{23.5.12}$$

The group $\mathrm{GL}_2(F)$ acts transitively on the left on the cosets $\mathrm{GL}_2(F)/(F^\times \mathrm{GL}_2(R))$ and we transport via the bijection (23.5.12) to an action on $\mathrm{Ver}(\mathcal{T})$.

We claim this action preserves the adjacency relation on $\mathcal{T}$: if $L \supseteq L'$ are adjacent, then by invariant factors we can choose a basis x_1, x_2 for L such that $x_1, \pi x_2$ is a basis for L', i.e.,

$$\beta' = \beta \begin{pmatrix} 1 & 0 \\ 0 & \pi \end{pmatrix}. \tag{23.5.13}$$

If $\alpha \in \mathrm{GL}_2(F)$, then multiplying (23.5.13) on the left by α shows that $\alpha L, \alpha L'$ are adjacent.

23.5.14. The tree $\mathcal{T}$ has a natural notion of distance d between two vertices, given by the length of the shortest path between them, giving each edge of $\mathcal{T}$ length 1. Consequently, we have a notion of distance $d(O, O')$ between every two maximal orders $O, O' \subseteq B$.

Lemma 23.5.15. *Let L, L' be lattices with bases x_1, x_2 and $\pi^f x_1, \pi^{e+f} x_2$, respectively. Then in the basis x_1, x_2, we have $O = \mathrm{End}_R(L) \simeq \mathrm{M}_2(R)$ and $O' = \mathrm{End}_R(L') \simeq \begin{pmatrix} R & \mathfrak{p}^{-e} \\ \mathfrak{p}^e & R \end{pmatrix}$, and $d(O, O') = e$.*

Proof. The statement on endomorphism rings comes from Example 10.5.2; we may suppose up to homothety that L' has basis $x_1, \pi^e x_2$; the maximal lattices as in 23.5.4 are given by $L_i = Rx_1 + \mathfrak{p}^i Rx_2$ with $i = 0, \dots, e$, so the distance is $d([L], [L']) = e$. □

23.5.16. Importantly, now, we turn to Eichler orders: they are the intersection of two unique maximal orders, and so correspond to a pair of vertices in $\mathcal{T}$, or equivalently a path. By Lemma 23.5.15, the standard Eichler order of level $\mathfrak{p}^e$ corresponds to a path of length e, and by transitivity the same is true of every Eichler order. The normalizer ϖ of an Eichler order 23.4.13 acts by swapping the two vertices. Each vertex of the path corresponds to the $e + 1$ possible maximal superorders.

In this way, the Bruhat–Tits tree provides a visual way to keep track of many calculations with Eichler orders.

Remark 23.5.17. The theory of Bruhat–Tits trees beautifully generalizes to become the theory of buildings, pioneered by Tits; see the survey by Tits [Tit79], as well as introductions by Abramenko–Brown [AB2008].

Exercises

Unless otherwise specified, let R be a Dedekind domain with $F = \mathrm{Frac}\, R$ and let $O \subseteq B$ be an R-order in a quaternion algebra B.

1. Let R be a DVR with maximal ideal $\mathfrak{p} = (\pi)$, and let
$$O = \begin{pmatrix} R & R \\ \mathfrak{p} & R \end{pmatrix}.$$
   (a) Suppose that $\alpha = \begin{pmatrix} x & y \\ \pi z & w \end{pmatrix} \in N_{B^\times}(O)$. After scaling, we may suppose that $x, y, z, w \in R$. By determinants, show that if $x, w \in R^\times$, then $\alpha \in O^\times$.
   (b) Compute
$$\alpha \varpi \overline{\alpha} = \begin{pmatrix} \pi(wy - xz) & x^2 - \pi y^2 \\ \pi(w^2 - \pi z^2) & -\pi(wy - xz) \end{pmatrix}.$$
   Show that $\pi \mid x$, and then $\pi \mid w$, whence $\alpha \in \varpi O$.

(c) Conclude that $N_{B^\times}(O)/(F^\times O^\times)$ is generated by $\varpi = \begin{pmatrix} 0 & 1 \\ \pi & 0 \end{pmatrix}$.

[See also the matrix proof by Eichler [Eic56a, Satz 5], which uses a normal form for one-sided ideals simplifying the above computation.]

▷ 2. Let $K \supseteq F$ be a field extension, let S be the integral closure of R in K, and let $\mathfrak{a}, \mathfrak{b} \subset F$ be fractional ideals of R. Show that $\mathfrak{a} = \mathfrak{b}$ if and only if $\mathfrak{a}S = \mathfrak{b}S$.

▷ 3. Extend Lemma 13.4.9, and show that if R is a DVR and O is an Eichler R-order then $\operatorname{nrd}(O^\times) = R^\times$.

▷ 4. Suppose R is local and O is a hereditary order. Show that if $\mu \in O$ has $\mu^2 = \pi$, then μ generates $N_{B^\times}(O)/(F^\times O^\times)$.

5. Suppose R is local and $O \subseteq B = \mathrm{M}_2(F)$ is the intersection of two maximal orders. Give another (independent) proof that O is isomorphic to a standard Eichler order which replaces matrix calculations in Proposition 23.4.3 with some representation theory as follows.

(a) Write $O \simeq \mathrm{M}_2(R) \cap O'$. Let e_{11} be the top-left matrix unit and let $I = \mathrm{M}_2(R)e_{11}$. Show $I' = O'e_{11}$ is an R-lattice in $V = \mathrm{M}_2(F)e_{11} \simeq F^2$.
(b) Use elementary divisors to show that there exists an R-basis x_1, x_2 of I such that $x_1, \pi^e x_2$ is an R-basis for I'.
(c) Show that the corresponding change of basis matrix $\alpha = \begin{pmatrix} 1 & 0 \\ 0 & \pi^e \end{pmatrix}$ has $I' = \alpha I$, and use this to identify O with the standard Eichler order of level $\mathfrak{p}^e$.

[See Brzezinski [Brz83a, Proposition 2.1].]

▷ 6. Let R be local and let O be the standard Eichler order of level $\mathfrak{p}^e$ for $e \geq 1$. Let $J = \operatorname{rad} O$. Show that

$$O_{\mathsf{L}}(J) = \begin{pmatrix} R & \mathfrak{p}^{-1} \\ \mathfrak{p}^{e-1} & R \end{pmatrix} = O_{\mathsf{R}}(J).$$

▷ 7. Prove Lemma 23.5.1. *[Hint: use direct matrix methods or the theory of invariant factors.]*

8. Let R be local and let O be a hereditary quaternion R-order. Show that $\operatorname{rad} O = [O, O]$ is the commutator (cf. Exercise 13.7) and that $\operatorname{diff} O = \operatorname{rad} O$.

9. Let R be local. Let $O, O' \subseteq B$ be maximal R-orders. Recall that O, O' are vertices in the Bruhat–Tits tree. Define the **distance** $\operatorname{dist}(O, O')$ to be the distance in the Bruhat–Tits tree between the respective vertices. Show that

$$[O : O \cap O'] = \operatorname{dist}(O, O') = [O' : O \cap O'].$$

10. Let R be local, and let O be an Eichler order of level $\mathfrak{p}^e$. Consider the graph whose vertices are R-superorders $O' \supseteq O$ in B and with a directed edge whenever the containment $O' \supsetneq O$ is proper and minimal. What does this graph look like? *[Hint: use the Bruhat–Tits tree; it helps to draw the Eichler orders of the same level at the same height.]*

This chapter is licensed under the terms of the Creative Commons Attribution-NonCommercial 4.0 International License (http://creativecommons.org/licenses/by-nc/4.0/), which permits any noncommercial use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license and indicate if changes were made.

The images or other third party material in this chapter are included in the chapter's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the chapter's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.

Chapter 24

Quaternion orders: second meeting

In this chapter, we continue our tour of quaternion orders with some more advanced species.

24.1 ▹ Advanced quaternion orders

Let B be a quaternion algebra over $\mathbb{Q}$ of discriminant $D = \operatorname{disc} B$ and let $O \subset B$ be an order with reduced discriminant $N = \operatorname{discrd}(O)$. Then $N = DM$ with $M \in \mathbb{Z}_{\geq 1}$.

24.1.1 (Gorenstein and Bass orders). Although Eichler orders may lose the property that all of its ideals are invertible, we may still insist on the invertibility of its dual. Recall (Definition 15.6.15) that the codifferent of an order is the lattice $\operatorname{codiff}(O) = O^\sharp$ obtained as the dual of the trace pairing over R. We say O is **Gorenstein** if $\operatorname{codiff}(O)$ is invertible, or equivalently (Corollary 16.8.7) every *sated* left or right fractional O-ideal (lattice $I \subseteq B$ with left or right order *equal* to O) is invertible. Hereditary orders are Gorenstein, since for a hereditary order every left or right fractional O-ideal (not necessarily sated) is invertible.

Being Gorenstein is a local property because invertibility is so. An Eichler order is Gorenstein, but there are Gorenstein orders that are not Eichler. An order is Gorenstein if and only if its associated ternary quadratic form is **primitive**, i.e. the greatest common divisor of its coefficients is 1, or equivalently its values generate $\mathbb{Z}$.

We say O is **Bass** if every superorder $O' \supseteq O$ (including $O' = O$) is Gorenstein. A Bass order is Gorenstein, but not always conversely. The fact that every superorder is Gorenstein reflects into good structural properties of a Bass order. Most importantly, a $\mathbb{Z}_p$-order O is Bass if and only if it contains either $\mathbb{Z}_p \times \mathbb{Z}_p$ or the ring of integers in a quadratic extension $K \supseteq \mathbb{Q}_p$ (these order are sometimes called *primitive*; we call them **basic**). This embedded subalgebra makes it possible to calculate explicitly with the order, with important applications to the arithmetic of modular forms, a topic we pursue in the final part of this book.

In summary, there is a chain of proper implications

$$\text{maximal} \Rightarrow \text{hereditary} \Rightarrow \text{Eichler} \Rightarrow \text{Bass} \Rightarrow \text{Gorenstein} \qquad (24.1.2)$$

© The Author(s) 2021

J. Voight, *Quaternion Algebras*, Graduate Texts in Mathematics 288,
https://doi.org/10.1007/978-3-030-56694-4_24

for orders $O \subseteq B$, and each of these implications is strict ($\not\Leftarrow$ in general).

Given an order O, we construct its **radical idealizer** as

$$O^\natural = O_\mathsf{L}(\operatorname{rad} O) = O_\mathsf{R}(\operatorname{rad} O).$$

We have $O \subseteq O^\natural$, and equality holds if and only if O is hereditary. Iterating, we obtain a canonically attached sequence of superorders:

$$O = O_0 \subsetneq O_1 = O^\natural \subsetneq \cdots \subsetneq O_s \tag{24.1.3}$$

where O_s is hereditary. A more refined classification of orders involves dissecting the chain (24.1.3) explicitly.

There is one final way of classifying orders that extends nicely to the noncommutative context, due to Brzezinski. By way of analogy, we recall that orders in a quadratic field are characterized by conductor. Let $K = \mathbb{Q}(\sqrt{d_K})$, let $\mathbb{Z}_K$ be the ring of integers of K, and let $d_K \in \mathbb{Z}$ be the discriminant of $\mathbb{Z}_K$. An order S in K is of the form $S = \mathbb{Z} + f\mathbb{Z}_K$, where $f \in \mathbb{Z}_{\geq 1}$ is the conductor of S (in its maximal order), and the discriminant of S is $d = f^2 d_K$. Even in classical considerations, these orders arise naturally when considering binary quadratic forms of nonfundamental discriminant.

Proposition 24.1.4. *Let B be a quaternion algebra over $\mathbb{Q}$ and let $O \subseteq B$ be an order. Then there exists a unique integer $f(O) \geq 1$ and Gorenstein order* $\operatorname{Gor}(O)$ *such that*

$$O = \mathbb{Z} + f(O)\operatorname{Gor}(O).$$

Two orders O, O' are isomorphic if and only if $f(O) = f(O')$ and $\operatorname{Gor}(O) \simeq \operatorname{Gor}(O')$.

The order $\operatorname{Gor}(O)$ is called the **Gorenstein saturation** of O, and we call $f(O)$ the **Gorenstein conductor** of the order (also sometimes called the **Brandt invariant**): an order is Gorenstein if and only if $f(O) = 1$, so this gives a ready supply of orders that are not Gorenstein. In particular, to classify all orders, via the operation of Gorenstein saturation, it is enough to classify the Gorenstein orders.

Remark 24.1.5. The first attempt to tame the zoo of quaternion orders was Eichler [Eic36, Satz 12], who classified what he called *primitive* (our *basic*) orders. Later this was generalized by Brzezinski [Brz83a, §5], who also clarified certain aspects [Brz90, §1]. A nice summary of facts about quaternion orders is given by Lemurell [Lem2011].

24.2 Gorenstein orders

In this section we define the well-behaved Gorenstein orders. See Remark 24.2.24 for more context on the class of Gorenstein rings.

Recall the definition of the codifferent (Definition 15.6.15):

$$\operatorname{codiff}(O) = O^\sharp := \{\alpha \in B : \operatorname{trd}(\alpha O) \subseteq R\} \subseteq B.$$

We have $\operatorname{codiff}(O)$ a two-sided sated O-ideal with $O \subseteq \operatorname{codiff}(O)$ (Lemma 15.6.16), and $\operatorname{disc}(O) = [\operatorname{codiff}(O) : O]_R$. We already saw in section 16.8 the importance of the following class of orders.

Definition 24.2.1. O is **Gorenstein** if $\operatorname{codiff}(O)$ is invertible.

24.2.2. The property of being a Gorenstein order is local (O is Gorenstein if and only if $O_{\mathfrak{p}}$ is Gorenstein for all primes $\mathfrak{p}$), since invertibility is a local property.

Proposition 24.2.3. *The following are equivalent:*

(i) *O is Gorenstein;*
(ii) $\operatorname{codiff}(O)$ *is projective as a O-bimodule;*
(iii) $O^\vee = \operatorname{Hom}_R(O, R)$ *is projective as an O-bimodule; and*
(iv) *All sated left or right fractional O-ideals are invertible as O-ideals.*

Proof. For the equivalence (i) $\Leftrightarrow$ (ii), because $\operatorname{codiff}(O)$ is sated it follows from Theorem 20.3.3 that O is Gorenstein if and only if $\operatorname{codiff}(O)$ is projective as an O-bimodule. For (ii) $\Leftrightarrow$ (iii), by Proposition 15.6.7, we have an isomorphism $\operatorname{codiff}(O) \simeq \operatorname{Hom}_R(O, R)$ of O-bimodules over R. Finally, (i) $\Leftrightarrow$ (iv) follows from Corollary 16.8.7. ◻

24.2.4. We call in for relief as well Main Theorem 20.3.9: the equivalent sided notions (on the left and right) in Proposition 24.2.3 are also all equivalent. In particular, a suitably defined notion of *left Gorenstein* or *right Gorenstein* would also be equivalent.

The Gorenstein condition can be detected on the level of norms as follows.

Lemma 24.2.5. *We have*

$$\operatorname{nrd}(\operatorname{codiff}(O))\operatorname{discrd}(O) \subseteq R,$$

and O is Gorenstein if and only if equality holds.

Proof. We refer to Proposition 16.4.3, and 16.4.5: we have

$$[O : O^\sharp]_R \supseteq \operatorname{Nm}_{B|F}(O^\sharp) = \operatorname{nrd}(O^\sharp)^2,$$

with equality if and only if $O^\sharp$ is locally principal. But by Lemma 15.6.17, we have $[O^\sharp : O]_R = \operatorname{disc}(O) = \operatorname{discrd}(O)^2$, and combining these gives the result. ◻

Our next main result connects the Gorenstein condition to a property of the corresponding ternary quadratic module. Let $Q\colon M \to L$ be a ternary quadratic module. We follow Gross–Lucianovic [GrLu2009, Propositions 6.1–6.2], and consider the odd Clifford bimodule.

Proposition 24.2.6. *Left multiplication gives a pairing*

$$\operatorname{Clf}^0(Q) \times \operatorname{Clf}^1(Q) \to \operatorname{Clf}^1(Q)/M \simeq \textstyle\bigwedge^3 M \otimes L^\vee$$

that induces an isomorphism

$$\operatorname{Hom}_R(\operatorname{Clf}^0(Q), \textstyle\bigwedge^3 M \otimes L^\vee) \xrightarrow{\sim} \operatorname{Clf}^1(Q) \tag{24.2.7}$$

of left $\operatorname{Clf}^0(Q)$*-modules.*

We have a similar argument on the right.

Remark 24.2.8. If R is a Dedekind domain, then choosing a pseudobasis for M we may write $M = \mathfrak{a}e_1 + \mathfrak{b}e_2 + \mathfrak{c}e_3$ for fractional ideals $\mathfrak{a}, \mathfrak{b}, \mathfrak{c}$, and without loss of generality may take $L = \mathfrak{l}$ a fractional ideal. Then $\bigwedge^3 M = \mathfrak{abc}(e_1 \wedge e_2 \wedge e_3) \simeq \mathfrak{abc}$ is just the Steinitz class of M, and $\bigwedge^3 M \otimes L^\vee \simeq \mathfrak{abc}\mathfrak{l}^{-1}$ as in (22.5.10). Restricting to this case is still quite illustrative.

Proof. By construction of the Clifford algebra, we have as R-modules that

$$\mathrm{Clf}^1(Q) \simeq M \oplus \left(\textstyle\bigwedge^3 M \otimes L^\vee\right),$$

so $\mathrm{Clf}^1(Q)/M \simeq \bigwedge^3 M \otimes L^\vee$. Multiplication in $\mathrm{Clf}(Q)$ induces a pairing that induces a homomorphism of $\mathrm{Clf}^0(Q)$-bimodules by associativity of multiplication in $\mathrm{Clf}(Q)$. To conclude that the pairing induces an isomorphism, we can argue locally and suppose that $M \simeq R^3$ with basis e_1, e_2, e_3 and $L \simeq R$ is the quadratic form

$$Q(x, y, z) = ax^2 + by^2 + cz^2 + uyz + vxz + wxy.$$

Then $1, e_2e_3, e_3e_1, e_1e_2$ is an R-basis for $\mathrm{Clf}^0(Q)$ and

$$\mathrm{Clf}^1(Q) = (Re_1 + Re_2 + Re_3) + Re_1e_2e_3$$

with $\mathrm{Clf}^1(Q)/M \simeq Re_1e_2e_3$.

We then compute the dual basis of $\mathrm{Clf}^1(Q)$ to that of $\mathrm{Clf}^0(Q)$ as

$$e_1e_2e_3 - ue_1 - ve_2 - we_3, e_1, e_2, e_3; \tag{24.2.9}$$

for example,

$$e_2e_3(e_1e_2e_3) = e_2(-e_1e_3 + v)e_2e_3 \equiv -e_2e_1e_3(e_2e_3) \pmod{M}$$

so

$$\begin{aligned} e_2e_3(e_1e_2e_3 - ue_1 - ve_2 - we_3) &\equiv e_1(e_2e_3)^2 - ue_1e_2e_3 \\ &\equiv e_1(ue_2e_3) - ue_1e_2e_3 = 0 \pmod{M}. \end{aligned}$$

The other products can be computed in a similarly direct fashion. □

Recall (Definition 9.7.14) that $Q\colon M \to L$ is *primitive* if $Q(M)$ generates L as an R-module.

Theorem 24.2.10. *O is Gorenstein if and only if its associated ternary quadratic module $Q = \psi_O$ is primitive.*

Proof. The statement is local, so we may suppose that R is a local domain with maximal ideal $\mathfrak{p}$, and that $M \simeq R^3$ and $L \simeq R$, so $\bigwedge^3 M \otimes L^\vee \simeq R$. Let $J = \mathrm{Clf}^1(Q)$ be the odd Clifford bimodule, thought of as a left O-module. By Proposition 24.2.6 (specifically, (24.2.7)), we have $\mathrm{Clf}^1(Q) \simeq \mathrm{Hom}_R(O, R)$ as left O-modules. By Proposition 24.2.3 and 24.2.4 (or repeating the argument on the right), we want to show that J is principal.

Suppose that Q is primitive. Let $\alpha = xe_1 + ye_2 + ze_3$. We then compute that

$$\alpha \begin{pmatrix} 1 \\ e_2e_3 \\ e_3e_1 \\ e_1e_2 \end{pmatrix} = \begin{pmatrix} x & 0 & cz+vx & -(by+uz) \\ y & -cz & vy & (ax+vz+wy) \\ z & (by+uz) & -(ax+wy) & 0 \\ 0 & x & y & z \end{pmatrix} \begin{pmatrix} e_1 \\ e_2 \\ e_3 \\ e_1e_2e_3 \end{pmatrix}$$

and the determinant of the matrix in the middle is precisely $Q(x,y,z)^2$. So $Q(\alpha) \in R^\times$ if and only if $\alpha O = J$.

Conversely, suppose that Q is not primitive. Then $Q \equiv 0 \pmod{\mathfrak{p}}$. If $\alpha = xe_1 + ye_2 + ze_3 + te_1e_2e_3$ with $x, y, z, t \in R$, then

$$\alpha e_2e_3 \equiv xe_1e_2e_3 \pmod{\mathfrak{p}}$$

and symmetrically with the other products, so αO has rank ≤ 2 over $R/\mathfrak{p}$, and it follows that $\alpha O \neq J$ for all α. □

Corollary 24.2.11. *Let R be a Dedekind domain. Then the associations*

$$\left\{ \begin{array}{c} \text{Nondegenerate } \textit{primitive} \text{ ternary} \\ \text{quadratic modules over } R \\ \text{up to twisted similarity} \end{array} \right\} \leftrightarrow \left\{ \begin{array}{c} \textit{Gorenstein} \text{ quaternion} \\ \text{orders over } R \\ \text{up to isomorphism} \end{array} \right\}$$
$$Q \mapsto \operatorname{Clf}^0(Q)$$
$$\operatorname{nrd}^\sharp(O) \leftarrow\!\shortmid O \tag{24.2.12}$$

are mutually inverse discriminant-preserving bijections that are also functorial with respect to R.

Proof. We restrict the bijection in Main Theorem 22.5.7 and apply Theorem 24.2.10. □

Remark 24.2.13. In view of Corollary 24.2.11, the issues in the correspondence with ternary quadratic forms for non-Gorenstein orders amounted to the failure to account for the codomain of the quadratic module: non-Gorenstein orders are obtained from quadratic modules $Q\colon M \to L$ where $Q(M) \subsetneq L$.

24.2.14. From 23.4.12 and Theorem 24.2.10, we conclude that every Eichler order is Gorenstein; a direct proof is given in Exercise 24.1.

Therefore, non-Gorenstein orders abound: indeed, any order corresponding to an imprimitive form will do. More generally, we construct a canonically associated Gorenstein order containing a given order as follows.

Proposition 24.2.15. *There exists a unique ideal $\mathfrak{f}(O) \subseteq R$ and unique Gorenstein order* $\operatorname{Gor}(O) \subseteq B$ *such that*

$$O = R + \mathfrak{f}(O)\operatorname{Gor}(O). \tag{24.2.16}$$

In fact, we have

$$\mathfrak{f}(O) = \operatorname{discrd}(O)\operatorname{nrd}(O^\sharp)$$
$$\operatorname{Gor}(O) = R + \operatorname{nrd}(O^\sharp)^{-1}O^\sharp O^\sharp.$$

Before we begin the proof, we make a definition and then consider the key ingredient of the proof: how rescaling the module affects the even Clifford algebra.

Definition 24.2.17. In (24.2.16), we call $\mathfrak{f}(O) \subseteq R$ the **conductor** and $\operatorname{Gor}(O)$ the **Gorenstein saturation** of O.

Remark 24.2.18. Brzezinski [Brz83a] writes $\mathfrak{b}(O)$ instead of $\mathfrak{f}(O)$ and calls it the **Brandt invariant**. Other authors call $\operatorname{Gor}(O)$ the *Gorenstein closure* of O, but this terminology may be confusing: $\operatorname{Gor}(O)$ is not necessarily the smallest Gorenstein order containing O (see Exercise 24.11).

24.2.19. Suppose that the ternary qudratic module $Q\colon M \to L$ corresponds to a quaternion R-order O. For a nonzero ideal $\mathfrak{a} \subseteq R$, we define $Q(\mathfrak{a}) : M \to \mathfrak{a}^{-1}L$ to be just Q but with values taken in $\mathfrak{a}^{-1} \supseteq L$. We claim that for all R-ideals $\mathfrak{a}$, we have under the correspondence

$$Q(\mathfrak{a})\colon M \to \mathfrak{a}^{-1}L \qquad O(\mathfrak{a}) := R + \mathfrak{a}O. \tag{24.2.20}$$

The fact that $\operatorname{Clf}^0(Q[\mathfrak{a}]) = R + \mathfrak{a}O$ is visible from the construction of the even Clifford algebra (22.2.3); it is also visible from the description (22.3.25) in Example 22.3.24. In the other direction, we have

$$\operatorname{nrd}^\sharp(O(\mathfrak{a}))\colon (O(\mathfrak{a})^\sharp)^0 = \mathfrak{a}^{-1}(O^\sharp)^0 \to \operatorname{discrd}(O(\mathfrak{a}))^{-1} = \mathfrak{a}^{-3}\operatorname{discrd}(O)^{-1}$$

and now we twist by $\mathfrak{a}$ to get

$$\mathfrak{a} \otimes \operatorname{nrd}(O(\mathfrak{a})^\sharp)\colon (O^\sharp)^0 \to \mathfrak{a}^{-1}\operatorname{discrd}(O)^{-1}. \tag{24.2.21}$$

Proof of Proposition 24.2.15. We argue using ternary quadratic modules: our proof amounts to replacing a potentially imprimitive form with a primitive form, following Theorem 24.2.10.

Let $Q = \psi_O : M \to L$ be the ternary quadratic module associated to O, well-defined up to twisted similarity. We may take $L = \mathfrak{l} \subseteq L \otimes_R F \simeq F$ and we do so for concreteness; accordingly, suppose $\mathfrak{l}$ is a fractional ideal of R. Then $Q(M) = \mathfrak{n} \subseteq \mathfrak{l}$ is a finitely generated nonzero R-submodule; since R is a Dedekind domain, $\mathfrak{n}$ is invertible. Let

$$\mathfrak{f} = \mathfrak{f}(O) := \mathfrak{n}\mathfrak{l}^{-1} \subseteq R.$$

Let $\operatorname{Gor}(Q) = Q(\mathfrak{f}) : M \to \mathfrak{n}$ be the primitive ternary quadratic module obtained by restricting the codomain. Then $\operatorname{Clf}^0(\operatorname{Gor}(Q))$ is a Gorenstein order by Theorem 24.2.10, and

$$\operatorname{Clf}^0(Q) = R + \mathfrak{f}(O)\operatorname{Clf}^0(\operatorname{Gor}(Q));$$

by (24.2.20), so we let $\operatorname{Gor}(O) := \operatorname{Clf}^0(\operatorname{Gor}(Q))$. Uniqueness follows directly from (24.2.20): if $O = R + \mathfrak{a}O'$ and O' is Gorenstein, then $Q(\mathfrak{a}^{-1}) : M \to \mathfrak{a}\mathfrak{l}$ is primitive and $Q(M) = \mathfrak{n} = \mathfrak{a}\mathfrak{l}$, thus $\mathfrak{a} = \mathfrak{f}(O)$ and $O' = \operatorname{Clf}^0(Q(\mathfrak{a})) = \operatorname{Gor}(O)$.

To prove the remaining statements, we recall Corollary 22.4.15 to get

$$O = R + \operatorname{discrd}(O)O^\sharp O^\sharp$$

in all cases, and the primitivity of

$$\operatorname{nrd}(O(\mathfrak{f}^{-1})^\sharp) \otimes \mathfrak{f}^{-1} \colon (O^\sharp)^0 \to \mathfrak{f}\operatorname{discrd}(O)^{-1}$$

as in (24.2.21) is equivalent to

$$\operatorname{nrd}((O^\sharp)^0) = \mathfrak{f}\operatorname{discrd}(O)^{-1}$$

i.e. $\mathfrak{f} = \operatorname{discrd}(O)\operatorname{nrd}((O^\sharp)^0)$. Finally, $\operatorname{nrd}(O^\sharp) = \operatorname{nrd}((O^\sharp)^0)$ is proven in Exercise 22.10. □

Lemma 24.2.22. *Let $O' \subseteq B$ be an R-order. Then $O \simeq O'$ as R-orders if and only if* $\operatorname{Gor}(O) \simeq \operatorname{Gor}(O')$ *and* $\mathfrak{f}(O) = \mathfrak{f}(O')$.

Proof. Immediate from the uniqueness claim in Proposition 24.2.15. □

The translation of the Gorenstein property in terms of primitivity of the ternary quadratic form makes it quite accessible. For example, we have the following result that shows that the Gorenstein condition is stable under base change.

Proposition 24.2.23. *Let $K \supseteq F$ be a finite extension and let S be the integral closure of R in K. Then O is a Gorenstein R-order if and only if $O \otimes_R S$ is a Gorenstein S-order.*

Proof. Let $Q \colon M \to L$ be the ternary quadratic module corresponding to O; denoting extension of scalars by subscripts, we have $Q_S : M_S \to L_S$ corresponding to O_S. We want to show that Q is primitive if and only if Q_S is primitive, which is to say $Q(M) = L$ if and only if $Q_S(M) = Q(M)_S = L_S$, and this statement is true as it holds for fractional R-ideals (Exercise 23.2). □

Remark 24.2.24. Gorenstein rings were introduced by Gorenstein [Gor52] in the context of plane curves (the results of his Ph.D. thesis); Bass [Bas62, Footnote 2] writes: "After writing this paper I discovered from Professor Serre that these rings have been encountered by Grothendieck, the latter having christened them 'Gorenstein rings.' They are described in his setting by the fact that a certain module of differentials is locally free of rank one." Bass [Bas63] gives a survey of (commutative) Gorenstein rings, noting their ubiquity; see also the later survey by Huneke [Hun99]. Gorenstein rings are truly abundant: they include coordinate rings of affine plane curves and curves with only double points as singularities, complete intersections, and integral group rings of finite groups.

The above-mentioned paper of Bass [Bas63] also gave rise to the class of eponymous orders in which every superorder is Gorenstein; over a complete DVR, these orders were completely classified (and related to hereditary and Gorenstein orders) by Drozd–Kirichenko–Roiter [DKR67]. Gorenstein orders in the context of quaternion algebras, were studied by Brzezinski [Brz82]. (Brzezinski [Brz87] also considers more general orders in which every lattice is locally principal: but for quaternion algebras, these are again the Gorenstein orders.)

24.3 Eichler symbol

Just as local quadratic extensions are classified as being either ramified, inert, or split, it is helpful to have a similar classification for quaternion orders.

We first work locally, and we suppose until further notice that R is local (as in 23.2.1).

24.3.1. The k-algebra $O/\operatorname{rad} O$ is semisimple and has a standard involution (since $\operatorname{rad} O$ is preserved by it), so by Example 7.4.9 this standard involution is nondegenerate; by classification (Main Theorem 4.4.1 and Theorem 6.4.1) we have one of three posibilities: $O/\operatorname{rad} O$ is either k, a separable quadratic k-algebra, or a quaternion algebra over k.

We give symbols to each of the possibilities in 24.3.1 as follows.

Definition 24.3.2. Let $J := \operatorname{rad} O$. We define the **Eichler symbol**

$$\left(\frac{O}{\mathfrak{p}}\right) := \begin{cases} *, & \text{if } O/J \text{ is a quaternion algebra;} \\ 1, & \text{if } O/J \simeq k \times k, \text{ and we say } O \text{ is } \textbf{residually split}; \\ 0, & \text{if } O/J \simeq k, \text{ and we say } O \text{ is } \textbf{residually ramified}; \text{ and} \\ -1, & \text{if } O/J \text{ is a (separable) quadratic field extension of } k, \text{ and} \\ & \text{we say } O \text{ is } \textbf{residually inert}. \end{cases}$$

For formatting reasons, we will also write $\left(\frac{O}{\mathfrak{p}}\right) = (O \mid \mathfrak{p})$. The similarity of the Eichler symbol to other quadratic-like symbols is intentional: because of the arguments to the symbol, it should not be confused with the others. If the reader finds this overloading of symbols unpleasant, they may wish to use the symbol $\varepsilon_{\mathfrak{p}}(O)$ instead.

24.3.3. Recall the definition of the discriminant quadratic form

$$\begin{aligned} \Delta : B &\to F \\ \Delta(\alpha) &= \operatorname{trd}(\alpha)^2 - 4\operatorname{nrd}(\alpha) \end{aligned} \tag{24.3.4}$$

that computes the discriminant of $F[\alpha] = F[x]/(x^2 - \operatorname{trd}(x)x + \operatorname{nrd}(x))$ in the basis $1, \alpha$. The form factors through a map $\Delta\colon B/F \to F$.

Suppose that $\#k < \infty$. For $a \in R$, let $\left(\frac{a}{\mathfrak{p}}\right)$ denote the generalized Kronecker symbol, defined to be $0, 1, -1$ according as if $F[x]/(x^2 - a)$ is ramified, split (isomorphic to $F \times F$), or inert. If $\operatorname{char} k \neq 2$, then $\left(\frac{a}{\mathfrak{p}}\right)$ is the Legendre symbol. We then have the following characterization (Exercise 24.4):

(a) $\left(\frac{O}{\mathfrak{p}}\right) = *$ if and only if $\left(\frac{\Delta(\alpha)}{\mathfrak{p}}\right)$ takes on all of the values $-1, 0, 1$ for $\alpha \in O$.

(b) $\left(\frac{O}{\mathfrak{p}}\right) = \epsilon$ if and only if $\left(\frac{\Delta(\alpha)}{\mathfrak{p}}\right)$ takes the values $\{0, \epsilon\}$ for $\alpha \in O$.

We now consider each of possible values of the Eichler symbol in turn.

24.3.5. We have $\left(\dfrac{O}{\mathfrak{p}}\right) = *$ if and only if $\operatorname{rad} O = \mathfrak{p}O$, by dimension considerations. If further $\#k < \infty$, then $\left(\dfrac{O}{\mathfrak{p}}\right) = *$ if and only if $O \simeq \mathrm{M}_2(R)$, since then the only quaternion algebra over k is $\mathrm{M}_2(k)$, and we can lift matrix units using Hensel's lemma.

Lemma 24.3.6. *The order O is Eichler if and only if O is maximal or residually split.*

Proof. If O is Eichler, then either O is maximal or $O/\operatorname{rad} O \simeq k \times k$ by (23.4.17). Conversely, a maximal order is an Eichler order by definition, so suppose O is residually split, i.e., $O/\operatorname{rad} O \simeq k \times k$. Then (as in Lemma 20.6.8), the nontrivial orthogonal idempotents of $k \times k$ lift to orthogonal idempotents $e_1, e_2 \in O$. Since $e_1 e_2 = 0$, immediately $B \simeq \mathrm{M}_2(F)$. Conjugating in B, we may suppose $e_1 = e_{11}$ and therefore $e_2 = e_{22}$, and then the result follows from Proposition 23.4.3.

For fun, here is a second proof. We have a decomposition $O = I_1 \oplus I_2 = Oe_1 \oplus Oe_2$. Thus $O = O_{\mathsf{L}}(I_1) \cap O_{\mathsf{L}}(I_2)$. Tensoring up to F, we get $B = (I_1)_F \oplus (I_2)_F$, so by dimensions we have $(I_1)_F \simeq (I_2)_F \simeq F^2$ the simple B-module, and each I_i is isomorphic to an R-lattice $M_i \subseteq F^2$. Thus $O_{\mathsf{L}}(I_i) \simeq \operatorname{End}_F(M_i) \simeq \mathrm{M}_2(R)$, and each $O_{\mathsf{L}}(I_i)$ is maximal, so O is the intersection of two maximal orders. $\square$

24.3.7. Residually inert orders in a division quaternion algebra B over $\mathbb{Q}_p$ were studied by Pizer [Piz76b], and he described them as follows. Let $K = \mathbb{Q}_{p^2}$ be the unramified extension of $\mathbb{Q}_p$. Then $K \hookrightarrow B$. Consider the left regular representation $B \to \mathrm{M}_2(K)$; it has image

$$\begin{pmatrix} z & w \\ p\overline{w} & \overline{z} \end{pmatrix}, \quad \text{with } z, w \in K \tag{24.3.8}$$

The valuation ring of B consists of those with $z, w \in \mathbb{Z}_{p^2}$ where $\mathbb{Z}_{p^2}$ is the valuation ring of K. Pizer then considers those orders with $z \in \mathbb{Z}_{p^2}$ and $p^r \mid w$. He [Piz80a, Remark 1.5, Proposition 1.6] connected the residually inert and residually split orders by noting the striking resemblance between (24.3.8) and the standard Eichler order, remarking:

> Thus O'_p and O_p [the Eichler order and the Pizer order] are both essentially subrings of $\begin{pmatrix} R & R \\ p^{2r+1} & R \end{pmatrix}$ fixed by certain (different!) Galois actions induced by the Galois group of $L/\mathbb{Q}_p$ and thus they can be viewed as twisted versions of each other. Hence O and O' are locally isomorphic at all primes $q \neq p$ while at p they are almost isomorphic. Thus it should not be too surprising that there are close connections between [them].

Pizer works explicitly and algorithmically [Piz80a] with residually inert orders, with applications to computing modular forms of certain nonsquarefree level.

24.3.9. We can also interpret the Eichler symbol in terms of the reduction of the associated ternary quadratic form Q.

We have $(\mathcal{O} \mid \mathfrak{p}) = *$ if and only if Q mod $\mathfrak{p}$ is nondegenerate, defining a smooth conic over k.

If $(\mathcal{O} \mid \mathfrak{p}) = 1$, then by Lemma 24.3.6, $\mathcal{O}$ is Eichler, and Q mod $\mathfrak{p} \sim Q(x, y) = xy$ is degenerate of rank 2 by 23.4.12, cutting out two intersecting lines over k; and conversely.

Suppose $(\mathcal{O} \mid \mathfrak{p}) = -1$. Let $i \in \mathcal{O}$ generate the quadratic field $\ell = \mathcal{O}/\operatorname{rad}\mathcal{O}$ over k, let $K = F(i)$ and let S be the integral closure of R in K. Then K is unramified over F, and $\mathfrak{p}S$ is the maximal ideal of S. Now the S-order $\mathcal{O}_S := \mathcal{O} \otimes_R S$ has $\mathcal{O}_S/\operatorname{rad}\mathcal{O}_S \simeq \ell \times \ell$, and $(\mathcal{O}_S \mid \mathfrak{p}S) = 1$. Therefore Q mod $\mathfrak{p}$ over ℓ is degenerate of rank 2, so the same is true over k, and since we are not in the previous case, it is defined by an irreducible quadratic polynomial (the norm form from ℓ to k). Therefore Q mod $\mathfrak{p}$ cuts out two lines defined over ℓ and conjugate under $\operatorname{Gal}(\ell \mid k)$. In particular, a residually inert order is Gorenstein.

The only possibilities that remain are that Q mod $\mathfrak{p}$ is identically zero or has rank 1 (defined by the square of a linear factor), and correspondingly cuts out the whole projective plane or a double line. These are the cases $(\mathcal{O} \mid \mathfrak{p}) = 0$.

24.3.10. It follows from 24.3.9 that the ternary quadratic form associated to a residually inert order is similar to

$$Q(x, y, z) = \pi^e x^2 + \operatorname{Nm}_{K|F}(z + yi)$$

just as in (23.2.7); we have e odd if and only if B is a division algebra and e even if and only if B is split.

The residually ramified orders do not admit such a simple classification; we will pursue them further in the coming sections.

Remark 24.3.11. Let R be a DVR, let $\mathcal{O}$ be a Gorenstein R-order, and let Q be a ternary quadratic form over R representing the similarity class associated to $\mathcal{O}$. Then Q is primitive and so defines an integral model $\mathcal{C}$ of a conic $C \subseteq \mathbb{P}^2$ over $F = \operatorname{Frac} R$. By discriminants, this conic has good reduction if and only if $\mathcal{O}$ is maximal. Moreover, we saw by direct calculation that this conic has the simplest kind of bad reduction—regular, with just one node over k—if and only if $\mathcal{O}$ is hereditary. This is no coincidence: in fact, $\mathcal{C}$ is normal if and only if $\mathcal{C}$ is Bass. See Brzezinski [Brz80] for more on the relationship between integral models of conics and quaternion orders and the follow-up work [Brz85] where the increasing sequence of Bass orders ending in an hereditary order corresponds to a sequence of elementary blowup transformations.

Another motivation to study the Eichler symbol is that it controls the structure of unit groups, as follows.

Lemma 24.3.12. *Let* $\#k = q$. *Then* $1 + \mathfrak{p}\mathcal{O} \subseteq \mathcal{O}^\times$, *and*

$$[\mathcal{O}^\times : 1 + \mathfrak{p}\mathcal{O}] = \begin{cases} q(q-1)^2(q+1), & \text{if } (\mathcal{O} \mid \mathfrak{p}) = *; \\ q^2(q-1)^2, & \text{if } (\mathcal{O} \mid \mathfrak{p}) = 1; \\ q^2(q^2-1), & \text{if } (\mathcal{O} \mid \mathfrak{p}) = -1; \\ q^3(q-1), & \text{if } (\mathcal{O} \mid \mathfrak{p}) = 0. \end{cases}$$

Proof. Since $J \subseteq \mathfrak{p}O$, if $\mu \in 1 + \mathfrak{p}O$ then $\mu - 1$ is topologically nilpotent, hence $\mu \in O^\times$.

Now the indices. If $(O \mid \mathfrak{p}) = *$, then we are computing the cardinality $\#\,\mathrm{GL}_2(k) = (q^2-1)(q^2-q)$. If $(O \mid \mathfrak{p}) = 1$, then $O/J \simeq k \times k$ and $\dim_k J/\mathfrak{p}O = 2$, so the index is $(q-1)^2q^2$. Similarly if $(O \mid \mathfrak{p}) = -1$, we get $(q^2-1)q^2$. If $(O \mid \mathfrak{p}) = 0$, then $O/J \simeq k$ and so the index is $(q-1)q^3$. □

We conclude with a global definition, and so we restore R to a Dedekind domain.

Definition 24.3.13. For a nonzero prime $\mathfrak{p} \subseteq R$, we define the **Eichler symbol at** $\mathfrak{p}$ to be

$$\left(\frac{O}{\mathfrak{p}}\right) := \left(\frac{O_\mathfrak{p}}{\mathfrak{p}R_\mathfrak{p}}\right),$$

i.e., the Eichler symbol of the completion at $\mathfrak{p}$.

We say O is **locally residually inert** if $\left(\dfrac{O}{\mathfrak{p}}\right) \in \{*, -1\}$ for all primes $\mathfrak{p}$.

The analogously defined *locally residually split* orders already have a name: they are the Eichler orders that are not maximal.

24.4 Chains of orders

We have a few more classes of orders to consider, but before we continue our tour we pause to consider an aspect of the more general classification: we seek to put every order in a chain of superorders, ending in a maximal order.

Suppose throughout this section that R is local.

Definition 24.4.1. The **radical idealizer** of O is

$$O^\natural := O_\mathsf{L}(\operatorname{rad} O) \cap O_\mathsf{R}(\operatorname{rad} O).$$

24.4.2. By Exercise 20.2, we have $O_\mathsf{L}(\operatorname{rad} O) = O_\mathsf{R}(\operatorname{rad} O) = O^\natural$, so the symmetric definition can be replaced by either order in the intersection.

24.4.3. We recall our motivation to study extremal orders 21.2.1: $O^\natural$ radically covers O, and by Proposition 21.2.3 we have $O^\natural = O$ if and only if O is extremal. By Theorem 21.5.1, O is extremal if and only if O is hereditary. Iterating, we have a canonically associated chain of orders

$$O = O_0 \subsetneq O_1 = O^\natural \subsetneq \cdots \subsetneq O_r \tag{24.4.4}$$

terminating in an order O_r that is hereditary. (By 23.3.1 and Proposition 23.4.3(iv), either O_r is maximal or O_r is contained in exactly two possible maximal orders.) We call O_r the **hereditary closure** of O.

24.4.5. Suppose $(O \mid \mathfrak{p}) = 1$, i.e., O is an Eichler order (Lemma 24.3.6). Suppose O has level $\mathfrak{p}^e$. Then $O^\natural$ is an Eichler order of level $\mathfrak{p}^{e-2}$ from (23.4.18)—it had to be Eichler of some level by Corollary 23.4.10. So the chain (24.4.4) is of length $\lfloor e/2 \rfloor$ with

quotients $\dim_k(O_i/O_{i+1}) = 2$. On the Bruhat–Tits tree, the Eichler order corresponds to a path of length e by 23.5.16, and $O^\natural$ is the path of length $e-2$ obtained by plucking away the vertices on the ends. (If desired, one can refine this chain by squeezing in an extra Eichler order in between each step.)

24.4.6. If $O = R + \mathfrak{p}O'$ for an order O', then $O^\natural = O'$, by Exercise 24.5.

In general, write $O = R + \mathfrak{p}^f \operatorname{Gor}(O)$ where $\mathfrak{p}^f = \mathfrak{f}(O)$ is the Gorenstein conductor of O and $\operatorname{Gor}(O)$ is the Gorenstein saturation as in Proposition 24.2.15. Then the chain of radical idealizers begins

$$O \subsetneq O_1 = R + \mathfrak{p}^{f-1} \operatorname{Gor}(O) \subsetneq \cdots \subsetneq O_f = \operatorname{Gor}(O).$$

For each i, we have $\dim_k(O_i/O_{i+1}) = 3$.

We next consider the chain of superorders over a (local) residually inert order.

Proposition 24.4.7. *Let O be a residually inert R-order. Then the following statements hold.*

(a) $\operatorname{rad} O = \operatorname{rad} O^\natural \cap O = \mathfrak{p}O^\natural$.

(b) *Suppose O is not maximal. Then $O^\natural$ is the unique minimal superorder of O. Moreover, $O^\natural$ is residually inert and we have $[O^\natural : O]_R = \mathfrak{p}^2$.*

Proof. We begin with the first part of (a), and we show $\operatorname{rad} O^\natural \cap O = \operatorname{rad} O$. As in 21.2.1, $O^\natural$ is a radical cover so $\operatorname{rad} O \subseteq \operatorname{rad} O^\natural \cap O$. But arguing as in the proof of Lemma 21.2.4, we know that $\operatorname{rad} O^\natural$ is topologically nilpotent as a O'-ideal and $\mathfrak{p}^r O' \subseteq \mathfrak{p}O$ for large r, so $\operatorname{rad} O^\natural \cap O$ is topologically nilpotent as a O-ideal, and $\operatorname{rad} O^\natural \cap O \subseteq \operatorname{rad} O$.

We therefore have a map

$$O/\operatorname{rad} O \hookrightarrow O^\natural/\operatorname{rad} O^\natural; \tag{24.4.8}$$

since $O/\operatorname{rad} O$ is a quadratic field, we must have $(O^\natural \mid \mathfrak{p}) = *, -1$, i.e., $O^\natural$ is either maximal or residually inert—and in the latter case, (24.4.8) is an isomorphism.

Let $i \in O$ generate $\ell = O/\operatorname{rad} O$ as a quadratic extension of k. Let $K = F(i)$, and let S be the integral closure of R in K. Then K is (separable and) unramified over F. We claim that $O_S = O \otimes_R S$ is residually split. Indeed, we have an isomorphism of k-algebras

$$(O/\mathfrak{p}O) \otimes_k \ell \xrightarrow{\sim} O_S/\mathfrak{p}O_S$$

and since ℓ is separable over k, an identification (Exercise 7.19)

$$\operatorname{rad}((O/\mathfrak{p}O) \otimes_k \ell) = \operatorname{rad}(O/\mathfrak{p}) \otimes \ell$$

giving $(\operatorname{rad} O)_S = \operatorname{rad}(O_S)$ and

$$O/\operatorname{rad} O \otimes_k \ell \simeq O_S/\operatorname{rad} O_S.$$

But $O/\operatorname{rad} O = \ell \otimes_k \ell \simeq \ell \times \ell$. This shows $(O_S \mid \mathfrak{p}S) = 1$.

To conclude, the statements that $\operatorname{rad} O^\natural \cap O = \mathfrak{p}O^\natural$ in (a) and that $[O^\natural : O]_R = \mathfrak{p}^2$ in (b) hold for O_S by (23.4.18), so they hold for O. Minimality follows from Proposition 24.4.12. ∎

24.4.9. As a consequence of Proposition 24.4.7, if O is a residually inert R-order with $\operatorname{discrd}(O) = \mathfrak{p}^e$, then the radical idealizer chain

$$O \subsetneq O_1 = O^{\natural} \subsetneq \cdots \subsetneq O_r$$

has length $r = \lfloor e/2 \rfloor$, with O_r *maximal*, and $\dim_k(O_i/O_{i+1}) = 2$ for all i; accordingly, the case e even occurs exactly when $B \simeq \mathrm{M}_2(F)$ and e odd occurs exactly when B is a division algebra.

We conclude the section showing that under certain hypotheses, the radical idealizer is a minimal (proper) superorder. The results are due to Drozd–Kirichenko–Roiter [DKR67, Propositions 1.3, 10.3]; we follow Curtis–Reiner [CR81, Exercises 37.5, 37.7].

Definition 24.4.10. Let $I' \subseteq I$ be left fractional O-ideals in B. We say I' is **(left) hypercharacteristic** in I if for every left O-module homomorphism $\phi\colon I' \to I$ we have $\phi(I') \subseteq I'$.

Lemma 24.4.11. *The map $O' \mapsto I' = (O')^{\sharp}$ gives an inclusion-reversing bijection from the set of R-superorders $O' \supseteq O$ to the set of right hypercharacteristic R-sublattices $I' \subseteq O^{\sharp}$.*

Proof. Inclusing-reversing follows from Lemma 15.6.2(a). By the proof of Lemma 17.3.3, we have a natural identification $\operatorname{Hom}_O(I', I) \simeq (I : I')_{\mathsf{R}}$ given by right multiplication.

We first show that if $O' \supseteq O$, then $(O')^{\sharp} \subseteq O$ is right hypercharacteristic; so we verify $(O^{\sharp} : (O')^{\sharp})_{\mathsf{R}} \subseteq (O')^{\sharp}$. If $(O')^{\sharp}\alpha \subseteq O^{\sharp}$, then

$$((O')^{\sharp}\alpha)^{\sharp} = \alpha^{-1}O' \supseteq (O^{\sharp})^{\sharp} = O$$

so $\alpha O \subseteq O'$, thus $\alpha = \alpha \cdot 1 \in O' = O_{\mathsf{R}}((O')^{\sharp})$. Conversely, given $I' \subseteq O^{\sharp}$ hypercharacteristic, we have $\alpha \in (I')^{\sharp}$ if and only if $\operatorname{trd}(I'\alpha) \subseteq R$ if and only if $\alpha \in (I' : O)_{\mathsf{R}} = (I' : I') = O_{\mathsf{R}}(I')$, so $(I')^{\sharp} = O_{\mathsf{R}}(I')$ is an R-order. □

Proposition 24.4.12. *Let R be local and let O be a Gorenstein R-order that is not maximal and such that O is indecomposable as a left O-module. Then there is a unique minimal R-superorder $O' \supsetneq O$ and $O' = O^{\natural}$.*

Proof. Since O is projective indecomposable as a O-module, and O is Gorenstein so $O^{\sharp}$ is projective, we must have $O^{\sharp} = O\alpha$ for some $\alpha \in B^{\times}$. Let $J = \operatorname{rad} O$. By Corollary 20.6.9, $JO^{\sharp}$ is the unique O-submodule of $O^{\sharp}$. If $JO^{\sharp}\beta = O^{\sharp}$ for some $\beta \in B^{\times}$, then $J = O\alpha\beta\alpha^{-1}$, so $O_{\mathsf{L}}(J) = O$ and O is extremal; but an extremal order that is indecomposable is already maximal, a contradiction. Therefore, if $JO^{\sharp}\beta \subseteq O^{\sharp}$ then the inclusion is strict, and by maximality we have $JO^{\sharp}\beta \subseteq JO^{\sharp}$; that is to say, $JO^{\sharp}$ is hypercharacteristic in $O^{\sharp}$. By Lemma 24.4.11, and inclusion-reversing, we see that

$$O' = (JO^{\sharp})^{\sharp} = (JO\alpha)^{\sharp} = \alpha^{-1}J^{\sharp}$$

is the unique minimal R-superorder, and $O' = O_{\mathsf{R}}(J^{\sharp}) = O_{\mathsf{L}}(J)$ by Proposition 15.6.6. □

24.4.13. Let R be local and O be a Gorenstein R-order. By Lemma 20.6.8, the condition that O is indecomposable is equivalent to the condition that $O/\operatorname{rad} O$ is simple as a k-algebra, i.e., $\left(\dfrac{O}{\mathfrak{p}}\right) \neq 1$ or equivalently, O is not Eichler. We considered the Eichler case in 24.4.5, so suppose further that O is not Eichler (and in particular, not maximal). Then $O^\natural \supseteq O$ is the minimal R-superorder over O by Proposition 24.4.12. If $O^\natural$ is not itself Gorenstein, then we may associate its Gorenstein saturation $\operatorname{Gor}(O^\natural)$. In this way, we obtain a canonical chain of Gorenstein superorders of O.

24.5 Bass and basic orders

Given the importance of the Gorenstein condition, we will want to give a name to the condition that every superorder is Gorenstein.

In this section, we restore our hypothesis that R is a Dedekind domain.

Definition 24.5.1. An order O is **Bass** if every order $O' \supseteq O$ is Gorenstein.

24.5.2. Since the Gorenstein condition is local by 24.2.2, the Bass condition is also local. Moreover, an Eichler (i.e., residually split) order is Bass, because Eichler orders are Gorenstein by 24.2.14, and every superorder of an Eichler order is Eichler (Corollary 23.4.10).

For the rest of this section, we investigate the local structure of Bass orders, and we suppose that R is local. We do not use the following proposition, but we state it for context.

Proposition 24.5.3. *Suppose R is local. Then the following are equivalent:*

(i) *Every O-ideal is generated by two elements;*
(ii) *O is Bass; and*
(iii) *Every O-lattice is isomorphic to a direct sum of O-ideals.*

Proof. See Drozd–Kirichenko–Roiter [DKR67, Propositions 12.1, 12.5] or Curtis–Reiner [CR81, §37]; the implications (i) $\Rightarrow$ (ii) $\Rightarrow$ (iii) always hold, and the implication (iii) $\Rightarrow$ (i) holds because B is a quaternion algebra. □

The residually inert orders, those with Eichler symbol $(O \mid \mathfrak{p}) = -1$, give a source of Bass orders, following Brzezinski [Brz83a, §3].

Proposition 24.5.4. *Let O be a residually inert R-order. Then O is Bass.*

Proof. We begin by arguing as in the proof of Proposition 24.4.7: letting $i \in O$ generate the field extension $O/P \supseteq k$, we then make a base extension to $K = F(i)$ where $i \in O$ generates $O/\operatorname{rad} O$: then O_S is residually split, i.e., O_S is Eichler.

We then appeal to Proposition 24.2.23: the Gorenstein condition is stable under base change. So for every superorder $O' \supseteq O$, we have a superorder $O'_S \supseteq O_S$ and so O'_S is Gorenstein by 24.5.2, so O' is Gorenstein. Therefore O is Bass. □

24.5.5. Combining 24.5.2 with Proposition 24.5.4, we see that if O is not a Bass order, then $\left(\dfrac{O}{\mathfrak{p}}\right) = 0$.

Another rich source of Bass orders are the basic orders.

Definition 24.5.6. We say O is **basic** if O contains a maximal R-order in a maximal commutative F-subalgebra $K \subseteq B$.

Remark 24.5.7. Local basic orders in a division quaternion algebra were studied by Hijikata–Pizer–Shemanske [HPS89b]; they gave the global orders O such that $O_{(\mathfrak{p})}$ is basic if $\mathfrak{p} \in \operatorname{Ram}(B)$ and Eichler if $\mathfrak{p} \notin \operatorname{Ram}(B)$ the name **special**. The remaining types of local basic orders (residually ramified and residually inert for the matrix ring) were studied by Brzezinski [Brz90], and further worked on by Jun [Jun97]. The role of the embedded maximal quadratic R-algebra S is that one can compute embedding numbers for them (see for example the epic work of Hijikata–Pizer–Shemanske [HPS89a]), and therefore compute explicitly with the trace formula.

Other authors use the term *primitive* instead of basic, but this quickly gets confusing as the word *primitive* is used for the ternary quadratic forms and the two notions do not coincide. The following propositions show that the sound of the word *basic* conveys the right meaning.

Proposition 24.5.8. *Suppose R is local. Then O is basic if and only if O is Bass.*

Proof. See Chari–Smertnig–Voight [CSV2019, Theorem 1.1]. To illustrate the proof, we explain here only the case where $2 \in R^\times$, following Brzezinski [Brz90, Proposition 1.11].

Let $Q\colon M \to R$ be the ternary quadratic form associated to O, a representative up to twisted similarity chosen so that Q is integral. We argue explicitly with a quadratic form (22.3.5) the multiplication table of the order as a Clifford algebra in a good basis (22.3.7).

First, suppose O is Bass. Then O is Gorenstein, so after rescaling we may suppose $Q \sim \langle -1, b, c\rangle$, and the multiplication table reads:

$$\begin{aligned} i^2 &= -bc & jk &= -\bar{i} \\ j^2 &= c & ki &= b\bar{j} \\ k^2 &= b & ij &= c\bar{k} \end{aligned} \tag{24.5.9}$$

If $\operatorname{ord}_{\mathfrak{p}}(b) \leq 1$ or $\operatorname{ord}_{\mathfrak{p}}(c) \leq 1$, then correspondingly $R[j]$ or $R[k]$ are maximal (again by valuation of discriminant), so suppose $\operatorname{ord}_{\mathfrak{p}}(b), \operatorname{ord}_{\mathfrak{p}}(c) \geq 2$. Consider the R-submodule O' generated by $1, i', j, k$ with $i' = i/\pi$. Then the integrality of the multiplication table remains intact so O' is an order, with new coefficients $a' = \pi$, $b' = b/\pi$, $c' = c/\pi$. The corresponding ternary quadratic form Q', by our hypotheses on valuations, is now imprimitive and O' is not Gorenstein, and this contradicts the fact that O is Bass.

Conversely, suppose O is basic. If the maximal R-order $S \subseteq O$ is in an unramified subalgebra $K = FS \subseteq B$, then it generates an unramified extension over the residue field and $(O \mid \mathfrak{p}) \neq 0$, and then by 24.5.5 we know O is Bass. Now at least one of the

products bc, ac, ab has valuation 1 because O contains a maximal R-order; without loss of generality we may take, after rescaling, $a = -1$ and $\operatorname{ord}_{\mathfrak{p}}(b) = 1 \le \operatorname{ord}_{\mathfrak{p}}(c)$ with $S = R[k]$. Therefore Q is primitive and O is Gorenstein.

We now compute that $\operatorname{rad} O = \langle \pi, i, j, k \rangle$ and so

$$O^{\natural} = O_{\mathsf{L}}(\operatorname{rad} O) = R + \mathfrak{p}^{-1} i + Rj + Rk$$

is the minimal superorder, and it is still basic. By minimality, to show that O is Bass it suffices to show that $O' = O^{\natural}$ is Bass. In our new parameters we have $a', b', c' = -\pi, b/\pi, c/\pi$, so $\operatorname{ord}_{\mathfrak{p}}(b') = 0$ and $\operatorname{ord}_{\mathfrak{p}}(c') = \operatorname{ord}_{\mathfrak{p}}(c) - 1$. Swapping i', j' interchanges a', b', and we are back in the original situation but with $\operatorname{ord}_{\mathfrak{p}}(c)$ reduced. By induction, we can continue in this way until $\operatorname{ord}_{\mathfrak{p}}(c) = 0$, when then $R[j]$ is a maximal order in an unramified extension, and we are done. □

Proposition 24.5.10. *Suppose $F = \operatorname{Frac}(R)$ is a number field. Then for the R-order O, the following are equivalent:*

(i) *O is basic;*
(ii) *$O_{(\mathfrak{p})}$ is basic for all primes $\mathfrak{p}$ of R; and*
(iii) *O is Bass.*

In other words, for orders in a quaternion algebra over a number field, being basic is a local property and it is equivalent to being Bass.

Proof. See Chari–Smertnig–Voight [CSV2019, Theorem 1.2], building on work of Eichler [Eic36, Satz 8] for the case $F = \mathbb{Q}$. □

Remark 24.5.11. One can similarly define basic orders for a general Dedekind domain R, but the preceding results are not known in this level of generality.

24.5.12. We established several other important features along the way in Proposition 24.5.8 that we now record. Suppose R is local with $2 \in R^{\times}$ and suppose that O is a residually ramified Bass (i.e., basic) order. Then the quadratic form associated to O is similar to $\langle -1, b, c \rangle$ with $\operatorname{ord}_{\mathfrak{p}}(b) = 1 \le \operatorname{ord}_{\mathfrak{p}}(c)$, and so the multiplication table (24.5.9) holds. The unique superorder $O^{\natural}$ has $[O^{\natural} : O]_R = \mathfrak{p}$ and associated ternary quadratic form $\langle -1, b, -c/b \rangle$ (see Exercise 24.13).

Corollary 24.5.13. *If $O^{\natural}$ is not hereditary, then* $\left(\dfrac{O^{\natural}}{\mathfrak{p}}\right) = \left(\dfrac{O}{\mathfrak{p}}\right)$.

Proof. Since $O^{\natural}$ is not hereditary, we cannot have O maximal, so $(O \mid \mathfrak{p}) = 1, 0, -1$. If $\left(\dfrac{O}{\mathfrak{p}}\right) = 1$, then O is Eichler, and so too are its superorders. If $\left(\dfrac{O}{\mathfrak{p}}\right) = -1$ and $O^{\natural}$ is not maximal, then $\left(\dfrac{O^{\natural}}{\mathfrak{p}}\right) = -1$ by Proposition 24.4.7(b). For the case $\left(\dfrac{O}{\mathfrak{p}}\right) = 0$, we appeal to 24.5.12. □

We can repackage what we have done for basic orders to give another description in terms of its hereditary closure.

Proposition 24.5.14. *Let O be a basic, nonhereditary R-order with* $\operatorname{discrd}(O) = \mathfrak{p}^n$. *Suppose $2 \in R^\times$, and let $S \subseteq O$ be a maximal R-order in the F-algebra K. Let J be the Jacobson radical of the hereditary closure of O. Then the following statements hold.*

(a) *Suppose O is residually inert. Then $O = S + J^m$ where $m = n/2, n-1$ according as B is ramified or split.*

(b) *Suppose O is residually ramified. Then $O = S + J^m$ where $m = n - 1$.*

Proof. The statement follows by induction using the explicit descriptions of these orders in 24.3.7, 24.5.12, and 24.5.12. See Brzezinski [Brz90, Proposition 1.12]. □

24.6 Tree of odd Bass orders

To conclude this chapter, we draw a picture of the containments of Bass orders.

Suppose R is local with finite residue field k and $2 \in R^\times$. We put together the radical idealizer chains in the residually split 24.4.5, residually inert 24.4.9, and residually ramified 24.5.12 cases. The resulting tree of Bass orders is shown in Figure 24.6.1.

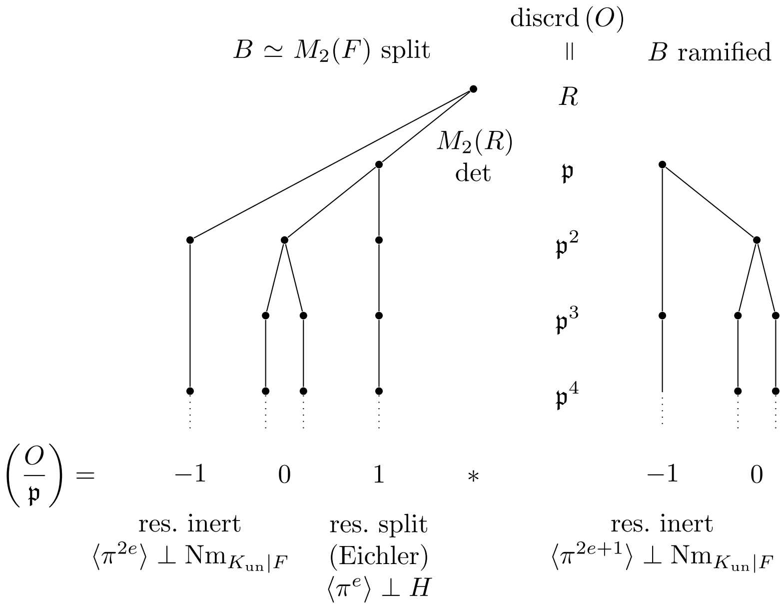


Figure 24.6.1: Tree of local Bass orders, odd characteristic residue field

Each vertex of this graph represents an isomorphism class of Bass order; there is an edge between two vertices if and only if there is a minimal containment between them. In each case, such a containment is given by the radical idealizer *except* when the order is residually split (in which case it hops by two, skipping the minimal superorder).

For the trees when 2 is a uniformizer in R, and many other explicit calculations, see Lemurell [Lem2011, §5], as well as Pacetti–Sirolli [PS2014, §5].

Exercises

Unless otherwise specified, let R be a Dedekind domain with $F = \operatorname{Frac} R$ and let $O \subseteq B$ be an R-order in a quaternion algebra B.

1. Show that every Eichler order O is Gorenstein by showing $O^\sharp$ is locally principal by direct computation.
2. Show that $\operatorname{codiff}(O)$ is right invertible if and only if
$$\operatorname{codiff}(O)(\operatorname{codiff}(O)^2)^\sharp = O$$
and $\operatorname{codiff}(O)$ is left invertible if and only if
$$(\operatorname{codiff}(O)^2)^\sharp \operatorname{codiff}(O) = O.$$
[Hint: show that $(\operatorname{codiff}(O)^{-1})^\sharp = \operatorname{codiff}(O)^2$.]

▷ 3. Suppose R is local. Show that if O is an R-order and B is a division algebra, then $\left(\dfrac{O}{\mathfrak{p}}\right) = 1, 0$.

▷ 4. Suppose R is local with finite residue field. Recall the discriminant quadratic form and the generalized Kronecker symbol 24.3.3.

   (a) Show that $\left(\dfrac{O}{\mathfrak{p}}\right) = *$ if and only if $\left(\dfrac{\Delta(\alpha)}{\mathfrak{p}}\right)$ takes on all of the values $-1, 0, 1$ for $\alpha \in O$.

   (b) For $\epsilon = -1, 0, 1$, show that $\left(\dfrac{O}{\mathfrak{p}}\right) = \epsilon$ if and only if $\left(\dfrac{\Delta(\alpha)}{\mathfrak{p}}\right)$ takes the values $\{0, \epsilon\}$ for $\alpha \in O$.

▷ 5. Let R be local and suppose $O = R + \mathfrak{p}O'$ for an order O'. Show that $\operatorname{rad} O = \mathfrak{p}O'$ and $O^\natural = O'$. *[Hint: Argue as in 23.4.15.]*

6. Let $Q\colon M \to L$ be a ternary quadratic module, and let $\mathfrak{a}$ be a fractional ideal. Write $\mathfrak{a}Q = Q \otimes \mathfrak{a}\colon M \otimes \mathfrak{a} \to L \otimes \mathfrak{a}^2$ for the twist. Show that there is a bijection $\operatorname{Cl} Q \leftrightarrow \operatorname{Cl} \mathfrak{a}Q$, and conclude that there is a bijection $\operatorname{Typ} O \leftrightarrow \operatorname{Typ} \operatorname{Gor}(O)$, where $\operatorname{Gor}(O)$ is the Gorenstein saturation.

7. (a) Show that if $\operatorname{discrd}(O)$ is cubefree (i.e., there is no prime $\mathfrak{p}$ of R such that $\mathfrak{p}^3 \mid \operatorname{discrd}(O)$) then O is a Bass order.

   (b) Show that there is a (local) Gorenstein order O with $\left(\dfrac{O}{\mathfrak{p}}\right) = 0$ that is not Bass with $\operatorname{discrd}(O) = \mathfrak{p}^4$.

8. Suppose that O is Gorenstein with $\left(\dfrac{O}{\mathfrak{p}}\right) \neq 1$. Let O' be the hereditary closure of O. Show that $N_{B^\times}(O) \leq N_{B^\times}(O')$, and further that equality holds when O is residually inert.

9. Let R be local and let O be a residually ramified quaternion R-order that is Gorenstein but not a Bass order. Let $O^\natural$ be the unique minimal order containing O. Show that $\mathfrak{f}(O^\natural) = \mathfrak{p}$. [See Brzezinski [Brz83a, Lemma 4.4].]

10. Let R be local with $2 \in R^\times$, and let $Q = \langle -1, b, c\rangle \colon R^3 \to R$ with $0 = \operatorname{ord}_{\mathfrak{p}}(b) < \operatorname{ord}_{\mathfrak{p}}(c)$. Let $O = \operatorname{Clf}^0(Q)$ be its even Clifford algebra. Show that

$$\left(\frac{O}{\mathfrak{p}}\right) = \left(\frac{b}{\mathfrak{p}}\right).$$

11. Let R be local, let $O = \mathrm{M}_2(R)$, let O_1 be the standard Eichler order of level $\mathfrak{p}^e$ and let $O_2 = R + \mathfrak{p}^e O$. Show that $O_2 \subsetneq O_1 \subsetneq O$, that $\operatorname{Gor}(O_2) = O$, but $\operatorname{Gor}(O_1) = O_1$; conclude that the Gorenstein saturation is *not* (necessarily) the smallest Gorenstein superorder.
12. Let R be local and let O be the standard Eichler order of level $\mathfrak{p}^e$ with $e \geq 2$. Show that O contains no integrally closed quadratic R-order that is a *domain* (even though O contains $R \times R$).

▹ 13. Let R be local with $2 \in R^\times$.

   (a) Show (using the proof of Proposition 24.5.8) that O is a local Bass order with $\left(\frac{O}{\mathfrak{p}}\right) = 0$ if and only if its corresponding ternary quadratic form is similar to $\langle -1, b, c\rangle$ with $\operatorname{ord}_{\mathfrak{p}}(b) = 1 \leq \operatorname{ord}_{\mathfrak{p}}(c)$.
   (b) In case (a), show that the minimal overorder O' corresponds to the (similarity class of) ternary quadratic form $\langle -1, b, -c/b\rangle$.

This chapter is licensed under the terms of the Creative Commons Attribution-NonCommercial 4.0 International License (http://creativecommons.org/licenses/by-nc/4.0/), which permits any noncommercial use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license and indicate if changes were made.

The images or other third party material in this chapter are included in the chapter's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the chapter's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.

Part III
Analysis

Chapter 25

The Eichler mass formula

In Part II of this text, we investigated arithmetic and algebraic properties of quaternion orders and ideals. Now in Part III, continuing this investigation we turn to the use of analytic methods. In this first introductory chapter, we restrict to a special case and consider zeta functions of quaternion orders over the rationals; as an application, we obtain a formula for the mass of the class set of a definite quaternion order.

25.1 ▹ Weighted class number formula

Gauss conjectured (in the language of binary quadratic forms) that there are finitely many imaginary quadratic orders of class number 1 [Gau86, Article 303]. Approaches to this problem involve beautiful and deep mathematics. Given that we want to prove some kind of lower bound for the class number (in terms of the discriminant of the order), it is natural to seek an analytic expression for it. The *analytic class number formula* of Dirichlet provides such an expression, turning the class number problem of Gauss into a (still very hard, but tractable) problem of estimation.

We recall section 14.1, which gives a classification of quaternion algebras over $\mathbb{Q}$, and sections 16.1 and 17.1, providing background on ideal classes in quaternion orders. With this motivation in hand, we are led to ask: what are the definite quaternion orders of class number 1? The method to prove Dirichlet's formula generalizes to definite quaternion orders as well, as pursued by Eichler in his *mass formula*. This chapter gives an overview of the Eichler mass formula in the simplest case for a maximal order in a definite quaternion algebra over $\mathbb{Q}$. (The reader who is already motivated and ready for action may consider skipping to the next chapter.)

Theorem 25.1.1 (Eichler mass formula over $\mathbb{Q}$,maximal orders). *Let B be a definite quaternion algebra over $\mathbb{Q}$ of discriminant D and let $\mathcal{O} \subset B$ be a maximal order. Then*

$$\sum_{[J] \in \operatorname{Cls} \mathcal{O}} \frac{1}{w_J} = \frac{\varphi(D)}{12}$$

where $w_J := \#\mathcal{O}_{\mathsf{L}}(J)^\times/\{\pm 1\}$ and $\varphi(D) := \#(\mathbb{Z}/D\mathbb{Z})^\times = \prod_{p \mid D}(p-1)$ is the Euler totient function.

© The Author(s) 2021

J. Voight, *Quaternion Algebras*, Graduate Texts in Mathematics 288,
https://doi.org/10.1007/978-3-030-56694-4_25

The Eichler mass formula does not quite give us a formula for the class number—rather, it gives us a formula for a "weighted" class number. That being said, we remark that $w_J \leq 24$ (see Theorem 11.5.14), and very often $w_J = 1$ (i.e., $O_{\mathsf{L}}(J) = \{\pm 1\}$. In order to convert the Eichler mass formula into a formula for the class number itself, one needs to understand the unit groups of left orders: this can be understood either as a problem in representation numbers of ternary quadratic forms or of embedding numbers of quadratic orders into quaternion orders, and we will take this subject up in earnest in Chapter 30.

Over $\mathbb{Q}$, the Eichler mass formula was first proven by Hey [Hey29, II, (80)], a Ph.D. student of Artin, along the same lines as the proof sketched below. This formula was also stated by Brandt [Bra28, §67]. We gradually warm up to this theorem by considering a broader analytic context. We see the analytic class number for an imaginary quadratic field as coming from the residue of its zeta function, and we then pursue a quaternionic generalization.

25.2 ▹ Imaginary quadratic class number formula

To introduce the circle of ideas, let $K := \mathbb{Q}(\sqrt{d})$ be a quadratic field of discriminant $d \in \mathbb{Z}$ and let $\mathbb{Z}_K$ be its ring of integers. We encode information about the field K by its *zeta function*.

25.2.1. Over $\mathbb{Q}$, we define the **Riemann zeta function**

$$\zeta(s) := \sum_{n=1}^{\infty} \frac{1}{n^s} \tag{25.2.2}$$

as the prototypical such function; this series converges for $\operatorname{Re} s > 1$, by the comparison test. By unique factorization, there is an **Euler product**

$$\zeta(s) = \prod_p \left(1 - \frac{1}{p^s}\right)^{-1} \tag{25.2.3}$$

where the product is over all primes p. The function $\zeta(s)$ can be meromorphically continued to the right half-plane $\operatorname{Re} s > 0$ using the fact that the sum

$$\zeta_2(s) = \sum_{n=1}^{\infty} \frac{(-1)^n}{n^s}$$

converges for $\operatorname{Re} s > 0$ and

$$\zeta(s) + \zeta_2(s) = 2^{1-s}\zeta(s)$$

so that

$$\zeta(s) = \frac{1}{2^{1-s} - 1}\zeta_2(s)$$

and the right-hand side makes sense for $\operatorname{Re} s > 0$ except for possible poles where $2^{1-s} = 1$. For real values of $s > 1$, we have

$$\frac{1}{s-1} = \int_1^\infty \frac{\mathrm{d}x}{x^s} \le \zeta(s) \le 1 + \int_1^\infty \frac{\mathrm{d}x}{x^s} = \frac{s}{s-1}$$

so

$$1 \le (s-1)\zeta(s) \le s;$$

therefore, as s approaches 1 from above, we have $\lim_{s\searrow 1}(s-1)\zeta(s) = 1$, so $\zeta(s)$ has a simple pole at $s = 1$ with residue

$$\operatorname{res}_{s=1} \zeta(s) = 1. \tag{25.2.4}$$

25.2.5. For the quadratic field K, modeled after (25.2.2) we define the **Dedekind zeta function** by

$$\zeta_K(s) := \sum_{\mathfrak{a}\subseteq\mathbb{Z}_K} \frac{1}{\mathsf{N}(\mathfrak{a})^s} \tag{25.2.6}$$

where $\mathsf{N}(\mathfrak{a}) := \#(\mathbb{Z}_K/\mathfrak{a})$ is the absolute norm, the sum is over all *nonzero* ideals of $\mathbb{Z}_K$, and the series is defined for $s \in \mathbb{C}$ with $\operatorname{Re} s > 1$. (We recall that $\mathsf{N}(\mathfrak{a})$ is the positive generator of $\operatorname{Nm}_{K|\mathbb{Q}}(\mathfrak{a})$, so we could equivalently work with the algebra norm, if desired.)

We can also write the Dedekind zeta function as a **Dirichlet series**

$$\zeta_K(s) = \sum_{n=1}^\infty \frac{a_n}{n^s} \tag{25.2.7}$$

where $a_n := \#\{\mathfrak{a} \subseteq \mathbb{Z}_K : \mathsf{N}(\mathfrak{a}) = n\}$ is the number of ideals in $\mathbb{Z}_K$ of norm $n \ge 1$. By unique factorization of ideals, we again have an Euler product expansion

$$\zeta_K(s) = \prod_{\mathfrak{p}} \left(1 - \frac{1}{\mathsf{N}(\mathfrak{p})^s}\right)^{-1}, \tag{25.2.8}$$

the product over all nonzero prime ideals $\mathfrak{p} \subset \mathbb{Z}_K$.

In order to introduce a formula that involves the class number, we group the ideals in (25.2.6) by their ideal class: for $[\mathfrak{b}] \in \operatorname{Cl}(K)$, we define the partial zeta function

$$\zeta_{K,[\mathfrak{b}]}(s) := \sum_{\substack{\mathfrak{a}\subseteq\mathbb{Z}_K \\ [\mathfrak{a}]=[\mathfrak{b}]}} \frac{1}{\mathsf{N}(\mathfrak{a})^s}$$

so that

$$\zeta_K(s) = \sum_{[\mathfrak{b}]\in\operatorname{Cl}(K)} \zeta_{K,[\mathfrak{b}]}(s). \tag{25.2.9}$$

In general, for $[\mathfrak{b}] \in \operatorname{Cl}(K)$, we have $[\mathfrak{a}] = [\mathfrak{b}]$ if and only if there exists $a \in K^\times$ such that $\mathfrak{a} = a\mathfrak{b}$, but since $\mathfrak{a} \subseteq \mathbb{Z}_K$, in fact

$$a \in \mathfrak{b}^{-1} = \{a \in \mathbb{Z}_K : a\mathfrak{b} \subseteq \mathbb{Z}_K\};$$

this gives a bijection

$$\{\mathfrak{a} \subseteq \mathbb{Z}_K : [\mathfrak{a}] = [\mathfrak{b}]\} \leftrightarrow \mathfrak{b}^{-1}/\mathbb{Z}_K^\times,$$

(since the generator of an ideal is unique up to units). Thus

$$\zeta_{K,[\mathfrak{b}]}(s) = \frac{1}{\mathrm{N}(\mathfrak{b})^s} \sum_{0 \neq a \in \mathfrak{b}^{-1}/\mathbb{Z}_K^\times} \frac{1}{\mathrm{N}(a)^s}. \tag{25.2.10}$$

for each class $[\mathfrak{b}] \in \mathrm{Cl}(K)$.

Everything we have done so far works equally as well for real as for imaginary quadratic fields. But to make sense of $\mathfrak{b}^{-1}/\mathbb{Z}_K^\times$ in the simplest case, we want $\mathbb{Z}_K^\times$ to be a finite group, which by Dirichlet's unit theorem means exactly that K is $\mathbb{Q}$ or an imaginary quadratic field. So from now on in this section, we suppose $K = \mathbb{Q}(\sqrt{d})$ with $d < 0$. Then $w := \#\mathbb{Z}_K^\times = 2$, except when $d = -3, -4$ where $w = 6, 4$, respectively.

Under this hypothesis, the sum (25.2.10) can be transformed into a sum over lattice points with the fixed factor w of overcounting. Before estimating the sum over reciprocal norms, we first estimate the count. Let $\Lambda \subset \mathbb{C}$ be a lattice. We can estimate the number of lattice points $\lambda \in \Lambda$ with $|\lambda| \le x$ by the ratio $\pi x^2/A$, where A is the area of a fundamental parallelogram P for Λ: roughly speaking, this says that we can tile a circle of radius x with approximately $\pi x^2/A$ parallelograms P.

More precisely, the following lemma holds.

Lemma 25.2.11. *Let $\Lambda \subset \mathbb{C}$ be a lattice with* $\mathrm{area}(\mathbb{C}/\Lambda) = A$. *Then there is a constant C such that for all $x > 1$,*

$$\left| \#\{\lambda \in \Lambda : |\lambda| \le x\} - \frac{\pi x^2}{A} \right| \le Cx.$$

We leave this lemma as an exercise (Exercise 25.3) in tiling a circle with radius x with fundamental parallelograms for the lattice Λ. With a bit of manipulation (Exercise 25.4), this lemma can be used to prove the analytic class number formula.

Theorem 25.2.12 (Analytic class number formula, imaginary quadratic field). *Let $K = \mathbb{Q}(\sqrt{d})$ be an imaginary quadratic field with discriminant $d < 0$. Then*

$$\mathrm{res}_{s=1}\, \zeta_K(s) = \frac{2\pi h}{w\sqrt{|d|}}$$

where h is the class number of K and w is the number of roots of unity in K.

This formula simplifies slightly if we cancel the pole at $s = 1$ with $\zeta(s)$, as follows. Like in the Dirichlet series, we can combine terms in (25.2.8) to get

$$\zeta_K(s) = \prod_p \prod_{\mathfrak{p} \mid p} \left(1 - \frac{1}{\mathrm{N}(\mathfrak{p})^s}\right)^{-1} = \prod_p L_p(p^{-s})^{-1}$$

where

$$L_p(T) := \begin{cases} (1-T)^2, & \text{if } p \text{ splits in } K; \\ 1 - T, & \text{if } p \text{ ramifies in } K; \text{ and} \\ 1 - T^2, & \text{if } p \text{ is inert in K.} \end{cases} \tag{25.2.13}$$

The condition of being split, ramified, or inert in K is recorded in a function:

$$\chi(p) := \chi_d(p) = \begin{cases} 1, & \text{if } p \text{ splits in } K; \\ 0, & \text{if } p \text{ ramifies in } K; \text{ and} \\ -1, & \text{if } p \text{ is inert in } K \end{cases} \tag{25.2.14}$$

for prime p and extended to all positive integers by multiplicativity. If $p \nmid d$ is an odd prime, then

$$\chi(p) = \left(\frac{d}{p}\right)$$

is the usual Legendre symbol, equal to 1 or -1 according as if d is a quadratic residue or not modulo p. Then in all cases, we have

$$L_p(T) = (1-T)(1-\chi(p)T).$$

Expanding the Euler product term-by-term and taking a limit, we conclude

$$\zeta_K(s) = \zeta(s)L(s,\chi) \tag{25.2.15}$$

where

$$L(s,\chi) := \prod_p \left(1 - \frac{\chi(p)}{p^s}\right)^{-1} = \sum_n \frac{\chi(n)}{n^s}. \tag{25.2.16}$$

The function $L(s,\chi)$ is in fact holomorphic for all $\operatorname{Re} s > 0$; this follows from the fact that the partial sums $\sum_{n \leq x} \chi(n)$ are bounded and the mean value theorem. So in particular the series

$$L(1,\chi) = 1 + \frac{\chi(2)}{2} + \frac{\chi(3)}{3} + \frac{\chi(4)}{4} + \dots$$

converges (slowly). Combining (25.2.15) with the analytic class number formula yields:

$$L(1,\chi) = \frac{2\pi h}{w\sqrt{|d|}} \neq 0. \tag{25.2.17}$$

For example, taking $d = -4$, so $\chi(2) = 0$ and $\chi(p) = (-1/p) = (-1)^{(p-1)/2}$,

$$\begin{aligned} L(1,\chi) &= 1 - \frac{1}{3} + \frac{1}{5} - \frac{1}{7} + \frac{1}{9} - \frac{1}{11} + \dots \\ &= \prod_{p \geq 3} \left(1 - \frac{(-1)^{(p-1)/2}}{p}\right)^{-1} = \frac{\pi}{4} = 0.7853\dots. \end{aligned}$$

Remark 25.2.18. The fact that $L(1,\chi) \neq 0$, and its generalization to complex characters χ, is the key ingredient to prove Dirichlet's theorem on primes in arithmetic progression (Theorem 14.2.9), used in the classification of quaternion algebras over $\mathbb{Q}$. The arguments to complete the proof are requested in Exercise 26.11.

Remark 25.2.19. To approach the class number problem of Gauss, we would then seek lower bounds on $L(1,\chi)$ in terms of the absolute discriminant $|d|$. Indeed, the history of class number problems is both long and beautiful. The problem of determining all positive definite binary quadratic forms with small class number was first posed by Gauss [Gau86, Article 303]. This problem was later seen to be equivalent to finding all imaginary quadratic fields of small class number (as in section 19.1). It would take almost 150 years of work, with important work of Heegner [Heeg52] and culminating in the results of Stark [Sta67] and Baker [Bak71], to determine those fields with class number 1: there are exactly nine, namely $d = -3, -4, -7, -8, -11, -19, -43, -67, -163$. See Goldfeld [Gol85] or Stark [Sta2007] for a history of this problem. For more specifically on the analytic class number formula for imaginary quadratic fields, see the survey by Weston [Wes] as well as the book by Serre [Ser73, Chapter VI].

25.3 ▹ Eichler mass formula: over the rationals

We are now prepared to consider the analogue of the analytic class number formula (Theorem 25.2.12) for quaternion orders: the Eichler mass formula, which is a *weighted* class number formula. We follow Eichler [Eic55-56, Eic56a], and in this section we give an overview with proofs omitted—a full development will be given in the next chapter, in more generality.

Let B be a quaternion algebra over $\mathbb{Q}$ of discriminant D and let $O \subset B$ be an order. We define the **zeta function** of O to be

$$\zeta_O(s) := \sum_{I \subseteq O} \frac{1}{\mathsf{N}(I)^s}, \tag{25.3.1}$$

where the sum over all invertible (nonzero, integral) right O-ideals and

$$\mathsf{N}(I) := [O : I] = \#(O/I) \in \mathbb{Z}_{>0}.$$

(By Main Theorem 16.1.3, we have $\mathsf{N}(I)$ the totally positive generator of $\operatorname{nrd}(I)^2$, so we could equivalently work with the reduced norm.)

Let a_n be the number of invertible right O-ideals of reduced norm $n > 0$ (with positive generator chosen, as usual). Then $\mathsf{N}(I) = \operatorname{Nm}(I) = \operatorname{nrd}(I)^2$ by 16.4.10, so

$$\zeta_O(s) = \sum_{n=1}^{\infty} \frac{a_n}{n^{2s}}. \tag{25.3.2}$$

To establish an Euler product for $\zeta_O(s)$, in due course we will give a kind of factorization formula for right ideals of O—but by necessity, writing an ideal as a compatible product will involve the entire set of orders connected to O! A direct consequence of the local-global dictionary for lattices (Theorem 9.4.9) is that

$$a_{mn} = a_m a_n \tag{25.3.3}$$

whenever m, n are coprime. Next, we will count the ideals of a given reduced norm $q = p^e$ a power of a prime: the answer will depend on the local structure of the order

O_p. Indeed, $\zeta_O(s)$ has an **Euler product**

$$\zeta_O(s) = \prod_p \zeta_{O,p}(p^{-s})^{-1} \tag{25.3.4}$$

with $\zeta_{O,p}(T) \in 1 + T\mathbb{Z}[T]$. In particular, $\zeta_O(s)$ only depends on the genus (local isomorphism classes) of O.

For simplicity, we first consider the case where O is a maximal order. Since there is a unique genus of maximal orders, the zeta function is independent of the choice of O and so we will write $\zeta_B(s) := \zeta_O(s)$ for O maximal. Then by a local count, we will show that

$$\zeta_{B,p}(T) = (1 - T^2) \cdot \begin{cases} 1, & \text{if } p \mid D; \\ 1 - pT^2, & \text{if } p \nmid D. \end{cases} \tag{25.3.5}$$

From (25.3.5),

$$\zeta_B(s) = \zeta(2s)\zeta(2s-1) \prod_{p \mid D} \left(1 - \frac{1}{p^{2s-1}}\right). \tag{25.3.6}$$

In particular, since $\zeta(s)$ has a simple pole at $s = 1$ with residue 1 and $\zeta(2) = \pi^2/6$ (Exercise 25.1),

$$\operatorname{res}_{s=1} \zeta_B(s) = \lim_{s \searrow 1} (s-1)\zeta_B(s) = \frac{\pi^2}{12} \prod_{p \mid D} \left(1 - \frac{1}{p}\right) \tag{25.3.7}$$

(We could also look to cancel the poles of $\zeta_B(s)$ in a similar way to define an L-function for B, holomorphic for $\operatorname{Re} s > 0$.)

Now we break up the sum (25.3.1) according to right ideal class:

$$\zeta_B(s) = \sum_{[J] \in \operatorname{Cls} O} \zeta_{B,[J]}(s)$$

where

$$\zeta_{B,[J]}(s) := \sum_{\substack{I \subseteq O \\ [I]=[J]}} \frac{1}{\mathsf{N}(I)^s}. \tag{25.3.8}$$

Since $[I] = [J]$ if and only if $I = \alpha J$ for some invertible $\alpha \in J^{-1}$, and $\mu J = J$ if and only if $\mu \in O_{\mathsf{L}}(J)^\times$, we conclude that

$$\zeta_{B,[J]}(s) = \frac{1}{\mathsf{N}(J)^s} \sum_{0 \neq \alpha \in J^{-1}/O_{\mathsf{L}}(J)^\times} \frac{1}{\mathsf{N}(\alpha)^s} \tag{25.3.9}$$

where the sum is taken over the nonzero elements $\alpha \in J^{-1}$ up to right multiplication by units $O_{\mathsf{L}}(J)^\times$ in the left order.

In order to proceed, we now suppose that B is definite (ramified at ∞) and hence that $\#O_{\mathsf{L}}(J)^\times < \infty$ (see Lemma 17.7.13); this is the analogue with the case of an imaginary quadratic field, and each J has the structure of a lattice in the Euclidean space $\mathbb{R}^4$ via the embedding

$$J \hookrightarrow B \hookrightarrow B \otimes_{\mathbb{Q}} \mathbb{R} \simeq \mathbb{H} \simeq \mathbb{R}^4. \tag{25.3.10}$$

Let $w_J = \#\mathcal{O}_\mathsf{L}(J)^\times/\{\pm 1\}$. We again argue by counting lattice points to prove the following proposition.

Proposition 25.3.11. *The function $\zeta_{B,[J]}(s)$ has a simple pole at $s = 1$ with residue*

$$\operatorname{res}_{s=1} \zeta_{B,[J]}(s) = \frac{\pi^2}{w_J D}.$$

Proof sketch. From a more general result (Theorem 26.2.12, proven in the next section and used to prove the analytic class number formula itself), we will show that

$$\operatorname{res}_{s=1} \zeta_{B,[J]}(s) = \frac{1}{2w_J \mathsf{N}(J)} \frac{\operatorname{vol}((\mathbb{R}^4)_{\leq 1})}{\operatorname{covol}(J)} \tag{25.3.12}$$

where under (25.3.10) we have

$$\operatorname{vol}((\mathbb{R}^4)_{\leq 1}) = \operatorname{vol}(\{x \in \mathbb{R}^4 : |x| \leq 1\}) = \frac{\pi^2}{2}$$

and

$$\operatorname{covol}(J) = \frac{\operatorname{covol}(\mathcal{O})}{\mathsf{N}(J)} = \frac{D/4}{\mathsf{N}(J)}.$$

Putting all of these facts together,

$$\operatorname{res}_{s=1} \zeta_{B,[J]}(s) = \frac{\pi^2}{4w_J \mathsf{N}(J)} \frac{4\mathsf{N}(J)}{D} = \frac{\pi^2}{w_J D}. \tag{25.3.13}$$

□

In particular the pole of each zeta function $\zeta_{B,[J]}(s)$ is *almost* independent of the class $[J]$, with the only relevant term being w_J the number of units.

Combining Proposition 25.3.13 with Proposition 25.3.11,

$$\operatorname{res}_{s=1} \zeta_B(s) = \frac{\pi^2}{D} \sum_{[J] \in \operatorname{Cls} \mathcal{O}} \frac{1}{w_J} = \frac{\pi^2}{12} \prod_{p \mid D} \left(1 - \frac{1}{p}\right) \tag{25.3.14}$$

and we conclude the following theorem.

Theorem 25.3.15 (Eichler mass formula, maximal orders). *Let B be a definite quaternion algebra over $\mathbb{Q}$ of discriminant D and let $\mathcal{O} \subset B$ be a maximal order. Then*

$$\sum_{[J] \in \operatorname{Cls} \mathcal{O}} \frac{1}{w_J} = \frac{\varphi(D)}{12}. \tag{25.3.16}$$

Remark 25.3.17. The Eichler mass formula is also very similar to the mass formula for the number of isomorphism classes of supersingular elliptic curves: this is no coincidence, and its origins will be explored in section 42.2.

To extend the Eichler mass formula to a more general class of orders, one only needs to replace the local calculation in 25.3.5 by a count of invertible ideals in the order. First we treat the important case of Eichler orders (see 23.1.3).

Theorem 25.3.18 (Eichler mass formula, Eichler orders over $\mathbb{Q}$). *Let $O \subset B$ be an Eichler order of level M in a definite quaternion algebra B of discriminant D. Then*

$$\sum_{[J]\in \mathrm{Cls}\, O} \frac{1}{w_J} = \frac{\varphi(D)\psi(M)}{12}$$

where

$$\psi(M) = \prod_{p^e \| M} (p^e + p^{e-1}) = M \prod_{p \mid M} \left(1 + \frac{1}{p}\right).$$

The most general formula is written in terms of the Eichler symbol 24.3. Just to recall two important cases: if $p \mid N = \mathrm{discrd}(O)$, then $\left(\dfrac{O}{p}\right) = -1$ if O_p is the maximal order in the division algebra B_p and $\left(\dfrac{O}{p}\right) = 1$ if O_p is an Eichler order.

Main Theorem 25.3.19 (Eichler mass formula, general case over $\mathbb{Q}$). *Let B be a definite quaternion algebra over $\mathbb{Q}$ and $O \subset B$ be an order with* $\mathrm{discrd}(O) = N$. *Then*

$$\sum_{[J]\in \mathrm{Cls}\, O} \frac{1}{w_J} = \frac{N}{12} \prod_{p \mid N} \lambda(O,p)$$

where

$$\lambda(O,p) = \frac{1-p^{-2}}{1-\left(\dfrac{O}{p}\right)p^{-1}} = \begin{cases} 1+1/p, & \textit{if } (O \mid p) = 1; \\ 1-1/p, & \textit{if } (O \mid p) = -1; \textit{ and} \\ 1-1/p^2, & \textit{if } (O \mid p) = 0. \end{cases} \tag{25.3.20}$$

Main Theorem 25.3.19 was proven by Brzezinski [Brz90, (4.6)] and more generally over number rings by Körner [Kör87, Theorem 1].

25.4 Class number one and type number one

The Eichler mass formula can be used to solve the class number 1 problem for definite quaternion orders over $\mathbb{Z}$, and in fact it is much easier than for imaginary quadratic fields! We begin with the case of maximal orders.

Theorem 25.4.1. *Let O be a maximal order in a definite quaternion algebra over $\mathbb{Q}$ of discriminant D. Then $\#\mathrm{Cls}\, O = 1$ if and only if $D = 2, 3, 5, 7, 13$.*

Proof. A calculation by hand; see Exercise 25.5. □

Remark 25.4.2. The primes $p = 2, 3, 5, 7, 13$ in Theorem 25.4.1 are also the primes p such that the modular curve $X_0(p)$ has genus 0. This is not a coincidence, and reflects a deep correspondence between classical and quaternionic modular forms (the *Eichler–Shimizu–Jacquet–Langlands correspondence*): see Remark 41.5.13.

The list of all definite quaternion orders (over $\mathbb{Z}$) of class number 1 was determined by Brzezinski [Brz95]. (Brzezinski mistakenly lists an order of class number 2, and so he counts 25, not 24; he corrects this later in a footnote [Brz98, Footnote 1].)

Theorem 25.4.3 (Brzezinski). *There are exactly* 24 *isomorphism classes of definite quaternion orders over* $\mathbb{Z}$ *with* $\#\operatorname{Cls} \mathcal{O} = 1$.

The list of orders with class number 1 is given in Table 25.4.4. We provide $N := \operatorname{discrd}(\mathcal{O})$, $D := \operatorname{disc}(B)$, we list the Eichler symbols $(\mathcal{O} \mid p)$ for the relevant primes $p \leq 13$, we say if the order is maximal, hereditary (but not maximal), Eichler/residually split or residually inert (but not hereditary), Bass (but not Eichler), or non-Gorenstein. This list includes the three norm Euclidean maximal orders (Exercise 25.6) of discriminant $D = 2, 3, 5$.

		$(\mathcal{O} \mid p)$							
N	D	2	3	5	7	11	13	Class	Q
2	2	−1	∗	∗	∗	∗	∗	maximal	$x^2 - xy - xz + y^2 + yz + z^2$
3	3	∗	−1	∗	∗	∗	∗	maximal	$x^2 - xy + y^2 + z^2$
4	2	0	∗	∗	∗	∗	∗	Bass	$x^2 + y^2 + z^2$
5	5	∗	∗	−1	∗	∗	∗	maximal	$x^2 - xy - xz + y^2 + yz + 2z^2$
6	2	−1	1	∗	∗	∗	∗	hereditary	$x^2 - xy + y^2 + 2z^2$
6	3	1	−1	∗	∗	∗	∗	hereditary	$x^2 + xz + y^2 - yz + 2z^2$
7	7	∗	∗	∗	−1	∗	∗	maximal	$x^2 - xz + y^2 + 2z^2$
8	2	−1	∗	∗	∗	∗	∗	Bass	$x^2 + xy - xz + y^2 - yz + 3z^2$
8	2	0	∗	∗	∗	∗	∗	Bass	$x^2 + y^2 + 2z^2$
10	2	−1	∗	1	∗	∗	∗	hereditary	$x^2 - xz + y^2 - yz + 3z^2$
10	5	1	∗	−1	∗	∗	∗	hereditary	$x^2 + xy + xz + 2y^2 + 2yz + 2z^2$
12	2	0	1	∗	∗	∗	∗	Bass	$x^2 + 2y^2 - 2yz + 2z^2$
12	3	−1	−1	∗	∗	∗	∗	residually inert	$x^2 + xy + y^2 + 4z^2$
12	3	1	−1	∗	∗	∗	∗	Eichler	$x^2 - xy + xz + 2y^2 - yz + 2z^2$
12	3	0	−1	∗	∗	∗	∗	Bass	$x^2 + y^2 + 3z^2$
13	13	∗	∗	∗	∗	∗	−1	maximal	$x^2 - xy + 2y^2 + yz + 2z^2$
16	2	0	∗	∗	∗	∗	∗	non-Gorenstein	$2(x^2 - xy - xz + y^2 + yz + z^2)$
16	2	0	∗	∗	∗	∗	∗	Bass	$x^2 + 2y^2 + 2z^2$
18	2	−1	−1	∗	∗	∗	∗	Bass	$x^2 + xz + y^2 - yz + 5z^2$
18	2	−1	1	∗	∗	∗	∗	Eichler	$x^2 + xz + 2y^2 + 2yz + 3z^2$
20	5	0	∗	−1	∗	∗	∗	Bass	$x^2 + 2y^2 - 2yz + 3z^2$
22	2	−1	∗	∗	∗	1	∗	hereditary	$x^2 + xz + 2y^2 + 3z^2$
24	3	0	−1	∗	∗	∗	∗	non-Gorenstein	$2(x^2 - xy + y^2 + z^2)$
28	7	−1	∗	∗	−1	∗	∗	Bass	$x^2 + xy - xz + 3y^2 - 2yz + 3z^2$

Table 25.4.4: Definite quaternion orders over $\mathbb{Z}$ with class number 1

Proof of Theorem 25.4.3. Suppose $\#\operatorname{Cls} \mathcal{O} = 1$. We apply the mass formula (Main Theorem 25.3.19). We note that $\lambda(\mathcal{O}, p) \geq 1 - 1/p$, in all three cases, so

$$1 \geq \frac{1}{w} = \frac{N}{12} \prod_{p \mid N} \lambda(\mathcal{O}, p) \geq \frac{N}{12} \prod_{p \mid N} \left(1 - \frac{1}{p}\right) = \frac{\varphi(N)}{12} \tag{25.4.5}$$

and therefore $\varphi(N) \leq 12$. By elementary number theory, this implies that

$$2 \leq N \leq 16 \text{ or } N = 18, 20, 21, 22, 24, 26, 28, 30, 36, 42.$$

This immediately gives a finite list of possibilities for the discriminant $D = \operatorname{disc} B \in \{2, 3, 5, 7, 11, 13, 30, 42\}$, as D must be a squarefree product of an odd number of primes.

By Exercise 17.3, if $O \subseteq O'$ then there is a natural surjection $\operatorname{Cls} O \to \operatorname{Cls} O'$, which is to say an order has at least as large a class number as any superorder. So we must have $\#\operatorname{Cls} O' = 1$ for a maximal order $O' \subseteq B$, and by Theorem 25.4.1, this then reduces us to $D = 2, 3, 5, 7, 13$. Because $\#\operatorname{Cls} O' = 1$, the maximal order O' is unique up to conjugation, and so fixing a choice of such a maximal order O' up to isomorphism we may suppose $O \subseteq O'$. But now the index $[O' : O] = M = D/N$ is explicitly given, and there are only finitely many suborders of bounded index; for each, we may compute representatives of the class set in a manner similar to the example in section 17.6. (We are aided further by deciding what assignment of Eichler symbols and unit orders would be necessary in each case.) □

There is similarly an interest in definite quaternion orders O of *type number* 1: these are the orders with the property that the "local-to-global principle applies for isomorphisms", i.e., if $O'_{\mathfrak{p}} \simeq O_{\mathfrak{p}}$ for all primes $\mathfrak{p}$ then $O' \simeq O$. If an order has class number 1 then it has type number 1, by Lemma 17.4.13, but one may have $\#\operatorname{Typ} O = 1$ but $\#\operatorname{Cls} O > 1$. Since an order has the same type number as its Gorenstein saturation, i.e. $\operatorname{Typ} O = \operatorname{Typ} \operatorname{Gor}(O)$, it suffices to classify the Gorenstein orders with this property.

By the bijection between ternary forms and quaternion orders, this is equivalently the problem of enumerating one-class genera of primitive ternary quadratic forms. The list was drawn up by Jagy–Kaplansky–Schiemann [JKS97] (with early work due to Watson [Wats75]), and has been independently confirmed by Lorch–Kirschmer [LK2013].

Theorem 25.4.6 (Watson, Jagy–Kaplansky–Schiemann, Lorch–Kirschmer). *There are exactly* 794 *primitive ternary quadratic forms of class number* 1, *corresponding to* 794 *Gorenstein quaternion orders of type number* 1. *The largest prime dividing a discriminant is* 23, *and the largest (reduced) discriminant is* $2^8 3^3 7^2 = 338688$. *There are exactly* 9 *corresponding to maximal quaternion orders: they have reduced discriminants*

$$D = 2, 3, 5, 7, 13, 30, 42, 70, 78.$$

Remark 25.4.7. The generalization of the class number 1 problem to quadratic forms of more variables was pursued by Watson, who showed that one-class genera do not exist in more than ten variables [Wats62]. Watson also tried to compile complete lists in low dimensions, followed by work of Hanke, and recently the complete list has been drawn up in at least 3 variables over $\mathbb{Q}$ by Lorch–Kirschmer [LK2013] and over totally real fields for maximal lattices by Kirschmer [Kir2014].

Exercises

1. A short and fun proof of the equality $\zeta(2) = \pi^2/6$ is due to Calabi [BCK93].

(a) Expand $(1-x^2y^2)^{-1}$ in a geometric series and integrate termwise over $S = [0,1]\times[0,1]$ to obtain

$$\iint_S (1-x^2y^2)^{-1}\,\mathrm{d}x\,\mathrm{d}y = 1 + \frac{1}{3^2} + \frac{1}{5^2} + \ldots = \left(1-\frac{1}{4}\right)\zeta(2).$$

(b) Show that the substitution

$$(x,y) := \left(\frac{\sin u}{\cos v}, \frac{\sin v}{\cos u}\right)$$

has Jacobian $1-x^2y^2$ and maps the open triangle

$$T := \{(u,v) : u,v > 0 \text{ and } u+v < \pi/2\}$$

bijectively to the interior of S.

(c) Conclude that

$$\iint_S (1-x^2y^2)^{-1}\,\mathrm{d}x\,\mathrm{d}y = \iint_T \mathrm{d}u\,\mathrm{d}v = \frac{\pi^2}{8}$$

and thus $\zeta(2) = \pi^2/6$.

2. In this exercise, we give a very jazzy proof that $\zeta(k) \in \mathbb{Q}\pi^k$ for all $k \in 2\mathbb{Z}_{\geq 1}$, due to Zagier [Zag94, p. 498].

(a) We start with $k = 4$. Define

$$f(m,n) := \frac{1}{mn^3} + \frac{1}{2m^2n^2} + \frac{1}{m^3n}$$

for $m,n \in \mathbb{Z}_{>0}$, and prove that

$$f(m,n) - f(m+n,n) - f(m,m+n) = \frac{1}{m^2n^2}.$$

(b) Prove

$$\zeta(2)^2 = \sum_{n=1}^{\infty} f(n,n) = \frac{5}{2}\zeta(4).$$

Conclude that $\zeta(4) = \pi^4/90$ using Exercise 25.1.

(c) In general, for $k \in 2\mathbb{Z}_{\geq 2}$, let

$$f(m,n) = \frac{1}{mn^{k-1}} + \frac{1}{2}\sum_{r=2}^{k-2}\frac{1}{m^r n^{k-r}} + \frac{1}{m^{k-1}n}$$

and check that

$$f(m,n) - f(m+n,n) - f(m,m+n) = \sum_{\substack{0<j<k \\ j \text{ even}}} \frac{1}{m^j n^{k-j}}.$$

Conclude in a similar way as in (b) that

$$\sum_{\substack{0<j<k \\ j \text{ even}}} \zeta(j)\zeta(k-j) = \frac{k+1}{2}\zeta(k)$$

so by induction $\zeta(k) \in \mathbb{Q}\pi^k$.

▹ 3. Prove Lemma 25.2.11 as follows.

(a) Let P be a fundamental parallelogram for Λ, and for $\lambda \in \Lambda$ let $P_\lambda := P + \lambda$. For $x > 1$, let

$$D(x) := \{z \in \mathbb{C} : |z| \leq x\}$$

and

$$\begin{aligned} N(x) &:= \#\{\lambda \in \Lambda : \lambda \in D(x)\} \\ N_P(x) &:= \#\{\lambda \in \Lambda : P_\lambda \subseteq D(x)\} \\ N_P^+(x) &:= \#\{\lambda \in \Lambda : P_\lambda \cap D(x) \neq \emptyset\}. \end{aligned}$$

Show that

$$N_P(x) \leq N(x) \leq N_P^+(x).$$

(b) Show that $N_P(x) \leq \pi x^2/A \leq N_P^+(x)$.

(c) Let l be the length of a long diagonal in P. Show that for all $\lambda \in \Lambda \cap D(x)$, we have $P_\lambda \subseteq D(x+l)$, so

$$N(x) \leq N_P(x+l) \leq \frac{\pi(x+l)^2}{A}.$$

Similarly, show that if $P_\lambda \cap D(x-l) \neq \emptyset$ then $P_\lambda \subseteq D(x)$ and $\lambda \in D(x)$, so

$$\frac{\pi(x-l)^2}{A} \leq N_P^+(x-l) \leq N(x).$$

(d) Conclude that Lemma 25.2.11 holds with $C := \pi(2l + l^2)/A$.

▹ 4. Using the previous exercise, we now prove the analytic class number formula (Theorem 25.2.12).

(a) Let $\mathfrak{b} \subset \mathbb{C}$ be a fractional ideal, and let

$$b_n := \#\{a \in \mathfrak{b}^{-1} : \mathrm{Nm}(a) = n\}.$$

Show that

$$\left| \sum_{n \leq x} b_n - \frac{\pi x}{A} \right| \leq C\sqrt{x}$$

for a constant C that does not depend on x and

$$A = \mathrm{Nm}(\mathfrak{b}^{-1})\frac{\sqrt{|d|}}{2}.$$

[Hint: Apply Lemma 25.2.11 to $\mathfrak{b}^{-1}$.]

(b) Consider the Dirichlet series

$$f(s) := \frac{1}{wN(\mathfrak{b})^s} \sum_{n=1}^{\infty} \left(b_n - \frac{\pi}{A}\right) \frac{1}{n^s}.$$

Show by the comparison test that $f(s)$ converges for all $\operatorname{Re} s > 1/2$.

(c) Show that

$$\operatorname{res}_{s=1} \zeta_K(s) = \lim_{s \searrow 1} (s-1)\zeta_{K,[\mathfrak{b}]}(s) = \frac{2\pi}{w\sqrt{|d|}}.$$

(d) Sum the residues over $[\mathfrak{b}] \in \operatorname{Cl}(K)$ to derive the theorem.

▹ 5. In this exercise, we prove Theorem 25.4.1: if O is a maximal order in a definite quaternion algebra over $\mathbb{Q}$ of discriminant D, then $\# \operatorname{Cls} O = 1$ if and only if $D = 2, 3, 5, 7, 13$. By the Eichler mass formula (Theorem 25.3.15), we have $\# \operatorname{Cls} O = 1$ if and only if

$$\frac{1}{w} = \frac{\varphi(D)}{12}$$

where $w = \# O/\{\pm 1\}$.

(a) Show (cf. 11.5.13) that if $D > 3$ then $w \leq 3$.

(b) Show that $\# \operatorname{Cls} O = 1$ for $D = 2, 3$. (The case $D = 2$ is the Hurwitz order and $D = 3$ is considered in Exercise 11.11. In fact, these orders are Euclidean with respect to the reduced norm: see the next exercise.)

(c) Show that if D is a squarefree positive integer with an odd number of prime factors and $\varphi(D)/12 \in \{1, 1/2, 1/3\}$, then $D \in \{5, 7, 13, 42\}$.

(d) Prove that $\# \operatorname{Cls} O = 1$ for $D = 5, 7, 13$ (cf. Exercise 17.10).

(e) Show that $\# \operatorname{Cls} O = 2$ for $D = 42$.

6. Let O be a definite quaternion order over $\mathbb{Z}$. If O is Euclidean, then $\# \operatorname{Cls} O = 1$, and we saw in 11.3.1 that the Hurwitz order O of discriminant $D = 2$ is Euclidean with respect to the reduced norm.

(a) Show that if O is norm Euclidean, then O is maximal.

(b) Show that O is Euclidean with respect to the reduced norm if and only if for all $\gamma \in B$, there exists $\mu \in O$ such that $\operatorname{nrd}(\gamma - \mu) < 1$.

(c) Show that if O is maximal, then O is norm Euclidean if and only if $D = 2, 3, 5$.

7. Generalizing the previous exercise, we may ask for the Euclidean ideal classes in maximal orders. We will show that there are no nonprincipal Euclidean two-sided ideal classes in maximal definite quaternion orders over $\mathbb{Z}$. [This exercise was suggested by Pete L. Clark.]

Let O be a maximal definite quaternion order over $\mathbb{Z}$ of discriminant D, and let I be a two-sided O-ideal. We say that I is **(norm) Euclidean** if for all $\gamma \in B$, there exists $\mu \in I$ such that $\operatorname{nrd}(\gamma - \mu) < \operatorname{nrd}(I)$.

(a) Show that if I is principal, then I is Euclidean if and only if O is Euclidean. In general, show that I is Euclidean if and only if αI is Euclidean for all $\alpha \in B^\times$, so we may ask if $[I]$ is Euclidean.

(b) Show that if $[I]$ is nontrivial and Euclidean, then $\#\operatorname{Cls} O = 2$ and $\operatorname{Pic}_{\mathbb{Z}}(O)$ is cyclic, generated by $[I]$. *[Hint: argue by induction on* $\operatorname{nrd}(J)$. *Then use the fact that* $\operatorname{Pic}(O)$ *is a group of exponent* 2.]

(c) Show that $\#\operatorname{Cls} O = 2$ if and only if $D = 11, 17, 19, 30, 42, 70, 78$.

(d) Show that $\operatorname{Pic}_{\mathbb{Z}}(O) = 1$ for $D = 11, 17, 19$, and for the remaining discriminants $D = 30, 42, 70, 78$ that the nontrivial class $[I]$ is *not* norm Euclidean.

This chapter is licensed under the terms of the Creative Commons Attribution-NonCommercial 4.0 International License (http://creativecommons.org/licenses/by-nc/4.0/), which permits any noncommercial use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license and indicate if changes were made.

The images or other third party material in this chapter are included in the chapter's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the chapter's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.

Chapter 26

Classical zeta functions

In this chapter, we prove the Eichler mass number for a definite quaternion order over a totally real field using classical analytic methods.

26.1 ▹ Eichler mass formula

In the previous chapter, we saw a sketch of how analytic methods with quaternionic zeta functions provide a weighted class number formula for a quaternion order in a definite quaternion algebra over $\mathbb{Q}$, analogous to the analytic class number formula of Dirichlet for a quadratic field. The main result of this section is then the generalization of the Eichler mass formula to a definite quaternion order over a totally real number field. In this section, we give the statement of this result.

26.1.1. Let F be a totally real number field of degree $n = [F : \mathbb{Q}]$, absolute discriminant d_F, and ring of integers $R := \mathbb{Z}_F$. Let h_F be the class number of F. Let

$$\zeta_F(s) := \sum_{\mathfrak{a} \subseteq R} \frac{1}{\mathsf{N}(\mathfrak{a})^s}$$

be the Dedekind zeta function of F, where $\mathsf{N}(\mathfrak{a}) = [R : \mathfrak{a}] \in \mathbb{Z}_{>0}$. Let B be a totally definite quaternion algebra over F of discriminant $\mathfrak{D}$. Let $O \subset B$ be an R-order with reduced discriminant $\operatorname{discrd}(O) = \mathfrak{N}$.

For a prime $\mathfrak{p} \mid \mathfrak{N}$ with $\mathsf{N}(\mathfrak{p}) = q$, let $\left(\dfrac{O}{\mathfrak{p}}\right) \in \{-1, 0, 1\}$ be the Eichler symbol (Definition 24.3.2), and let

$$\lambda(O, \mathfrak{p}) := \frac{1 - \mathsf{N}(\mathfrak{p})^{-2}}{1 - \left(\dfrac{O}{\mathfrak{p}}\right)\mathsf{N}(\mathfrak{p})^{-1}} = \begin{cases} 1 + 1/q, & \text{if } (O \mid p) = 1; \\ 1 - 1/q, & \text{if } (O \mid p) = -1; \\ 1 - 1/q^2, & \text{if } (O \mid p) = 0. \end{cases} \tag{26.1.2}$$

26.1.3. We saw in Lemma 17.7.13 that for each definite order O, the group O^1 of units of reduced norm 1 is a finite group; we will see in Lemma 26.5.1 that further the group $O^\times/R^\times$ is finite. For a right O-ideal J, the automorphism group of J (as a right

© The Author(s) 2021

J. Voight, *Quaternion Algebras*, Graduate Texts in Mathematics 288,
https://doi.org/10.1007/978-3-030-56694-4_26

O-module) consists of right multiplication maps by elements $\mu \in B^\times$ with $\mu J = J$, i.e., $\mu \in O_{\mathsf{L}}(J)^\times$.

It was already evident in the Eichler mass formula (and remains a general principle in mathematics) that one often gets a better count of objects when they are weighted by the inverse size of the automorphism group, so we weight a right ideal class $[J]$ by $[O_{\mathsf{L}}(J)^\times : R^\times]^{-1}$ and make the following definition of a *weighted* class number.

Definition 26.1.4. Define the **mass** of $\operatorname{Cls} O$ to be

$$\operatorname{mass}(\operatorname{Cls} O) = \sum_{[J]\in \operatorname{Cls} O} [O_{\mathsf{L}}(J)^\times : R^\times]^{-1}.$$

Main Theorem 26.1.5 (Eichler mass formula). *With notation as in 26.1.1, we have*

$$\operatorname{mass}(\operatorname{Cls} O) = \frac{2\zeta_F(2)}{(2\pi)^{2n}} d_F^{3/2} h_F \mathsf{N}(\mathfrak{N}) \prod_{\mathfrak{p} \mid \mathfrak{N}} \lambda(O, \mathfrak{p}). \tag{26.1.6}$$

26.1.7. The functional equation for the Dedekind zeta function relates s to $1-s$, giving an alternative way of writing (26.1.6) as

$$\frac{2\zeta_F(2)}{(2\pi)^{2n}} d_F^{3/2} = \frac{|\zeta_F(-1)|}{2^{n-1}}. \tag{26.1.8}$$

We notice that the Eichler mass formula then implies that $\zeta_F(-1) \in \mathbb{Q}$.

Remark 26.1.9. More generally, the rationality of the values $\zeta_F(-n)$ with $n \in \mathbb{Z}_{>0}$ is a theorem of Siegel [Sie69] and Deligne–Ribet [DR80].

Remark 26.1.10. The weighting in the mass is what makes Main Theorem 26.1.5 so simple. In the (unlikely) situation where $w_J = w_O$ is independent of J, we would have a formula for the class number, but more generally we will need to take account of unit groups by computing embedding numbers of cyclotomic quadratic orders: we will do this in Chapter 30.

We now make the formula (26.1.6) a bit more explicit for the case of Eichler orders.

26.1.11. Let O be an Eichler order of level $\mathfrak{M}$, so that $\mathfrak{N} = \mathfrak{D}\mathfrak{M}$ with $\mathfrak{D}, \mathfrak{M}$ coprime. Then

$$\left(\frac{O}{\mathfrak{p}}\right) = \begin{cases} -1, & \text{if } \mathfrak{p} \mid \mathfrak{D}; \\ 1, & \text{if } \mathfrak{p} \mid \mathfrak{M}; \\ *, & \text{if } \mathfrak{p} \nmid \mathfrak{N}. \end{cases}$$

Accordingly, we define the generalized Euler φ-function and Dedekind ψ-function by

$$\varphi(\mathfrak{D}) := \prod_{\mathfrak{p}\mid\mathfrak{D}} (\mathsf{N}(\mathfrak{p}) - 1) = \mathsf{N}(\mathfrak{D}) \prod_{\mathfrak{p}\mid\mathfrak{D}} \left(1 - \frac{1}{\mathsf{N}(\mathfrak{p})}\right)$$

$$\psi(\mathfrak{M}) := \prod_{\mathfrak{p}^e \parallel \mathfrak{M}} \mathsf{N}(\mathfrak{p})^{e-1}(\mathsf{N}(\mathfrak{p}) + 1) = \mathsf{N}(\mathfrak{M}) \prod_{\mathfrak{p}\mid\mathfrak{M}} \left(1 + \frac{1}{\mathsf{N}(\mathfrak{p})}\right)$$

(recalling $\mathfrak{D}$ is squarefree, with the natural extension $\varphi(\mathfrak{D}) = \#(R/\mathfrak{D})^\times$ for all $\mathfrak{D}$). The ψ-function computes a unit index: see Lemma 26.6.7.

The Eichler mass formula (Main Theorem 26.1.5) for Eichler orders then reads as follows.

Theorem 26.1.12 (Eichler mass formula, Eichler orders). *With notation as in 26.1.11, we have*

$$\operatorname{mass}(\operatorname{Cls} O) = \frac{2\zeta_F(2)}{(2\pi)^{2n}} d_F^{3/2} h_F \varphi(\mathfrak{D})\psi(\mathfrak{M}). \tag{26.1.13}$$

Remark 26.1.14. The Eichler mass formula in the form (26.1.13) for maximal orders was proven by Eichler (working over a general totally real field) using the techniques in this chapter [Eic38b, Satz 1], and was extended to squarefree level $\mathfrak{N}$ (i.e., hereditary orders) again by Eichler [Eic56a, §4]. This was extended by Brzezinski [Brz90, (4.6)] to a general formula over $\mathbb{Q}$ and by Körner [Kör87, Theorem 1], using idelic methods.

The classical method to prove Main Theorem 26.1.5 is similar to the one we sketched over $\mathbb{Q}$ in chapter 25, with some added technicalities of working over a number field. We follow this approach, first proving the formula when O is a maximal order, and then deducing the general case. We will return in chapter 29 and reconsider the Eichler mass formula from an idelic point of view, thinking of it as a special case of a volume formula (for a finite set of "quotient points"). It is hoped that this chapter will serve to show both the power and limits of classical methods before we build upon them using idelic methods.

26.2 Analytic class number formula

In this section, in preparation for the quaternionic case we briefly review what we need from the analytic class number formula for a number field F. References for this material include Borevich–Shafarevich [BS66, Chapter 5], Lang [Lang94, Chapter VI], and Neukirch [Neu99, Chapter VII].

We begin by setting some notation that will be used throughout the rest of this chapter. Let F be a number field of degree $n := [F : \mathbb{Q}]$ with r real places and c complex places, so that $n = r + 2c$. Let $R := \mathbb{Z}_F$ be the ring of integers in F, and let d_F be the discriminant of F. Let w_F be the number of roots of unity in F, let $h_F := \#\operatorname{Cl} R$ be the class number of F, and let Reg_F be the regulator of F (the covolume of $R^\times$ under the Minkowski embedding).

Define the **Dedekind zeta function** for $s \in \mathbb{C}$ with $\operatorname{Re}(s) > 1$ by

$$\zeta_F(s) := \sum_{\mathfrak{a} \subseteq R} \frac{1}{\mathsf{N}(\mathfrak{a})^s}$$

where the sum is over all nonzero ideals of R and $\mathsf{N}(\mathfrak{a}) = \#(R/\mathfrak{a}) = [R : \mathfrak{a}]$ is the absolute norm; we have $\mathsf{N}(\mathfrak{a}) = \operatorname{Nm}(\mathfrak{a})$ with norm taken from F to $\mathbb{Q}$ and positive generator chosen.

26.2.1. The Dedekind zeta function converges for $\operatorname{Re} s > 1$ and has an Euler product

$$\zeta_F(s) = \prod_{\mathfrak{p}} \left(1 - \frac{1}{\mathsf{N}(\mathfrak{p})^s}\right)^{-1} \tag{26.2.2}$$

where the product is over all nonzero primes of R—this follows formally from unique factorization of ideals after one shows that the pruned product converges.

The Dedekind zeta function has properties analogous to the Riemann zeta function, which is the special case $F = \mathbb{Q}$. In particular, we can extend $\zeta_F(s)$ in a manner analogous to 25.2.1 to $\operatorname{Re} s > 0$. For $a \in \mathbb{C}$ we write $\zeta_F^*(a)$ for the leading coefficient in the Laurent series expansion for ζ_F at $s = a$.

Theorem 26.2.3 (Analytic class number formula). *$\zeta_F(s)$ has analytic continuation to* $\operatorname{Re} s > 0$*, with a simple pole at $s = 1$ having residue*

$$\zeta_F^*(1) = \lim_{s \to 1}(s-1)\zeta_F(s) = \frac{2^r(2\pi)^c}{w_F\sqrt{|d_F|}} h_F \operatorname{Reg}_F. \tag{26.2.4}$$

Remark 26.2.5. The formula (26.2.4) is known as Dirichlet's **analytic class number formula** (even though the original form of Dirichlet's theorem concerned quadratic forms rather than classes of ideals, so is closer to Theorem 25.2.12).

Example 26.2.6. When F is an imaginary quadratic field ($r = 0$ and $c = 1$) we have $\operatorname{Reg}_F = 1$ and Theorem 26.2.3 is Theorem 25.2.12.

Before we finish this section, we review a few ingredients from the proof of the analytic class number formula (26.2.4) to set up the Eichler mass formula.

26.2.7. We first write the Dedekind zeta function as a sum over ideals in a given ideal class $[\mathfrak{b}] \in \operatorname{Cl}(R)$: we define the partial zeta function

$$\zeta_{F,[\mathfrak{b}]}(s) := \sum_{\substack{\mathfrak{a} \subseteq R \\ [\mathfrak{a}]=[\mathfrak{b}]}} \frac{1}{\mathsf{N}(\mathfrak{a})^s} \tag{26.2.8}$$

convergent for $\operatorname{Re} s > 1$ by comparison to the harmonic series, so that

$$\zeta_F(s) = \sum_{[\mathfrak{b}] \in \operatorname{Cl} R} \zeta_{F,[\mathfrak{b}]}(s). \tag{26.2.9}$$

Now note that $[\mathfrak{a}] = [\mathfrak{b}]$ if and only if $\mathfrak{a} = a\mathfrak{b}$ for some nonzero

$$a \in \mathfrak{b}^{-1} = \{x \in F : x\mathfrak{b} \subseteq R\},$$

so there is a bijection between nonzero ideals $\mathfrak{a} \subseteq R$ such that $[\mathfrak{a}] = [\mathfrak{b}]$ and the set of nonzero elements in $\mathfrak{b}^{-1}/R^\times$. So

$$\zeta_{F,[\mathfrak{b}]}(s) = \frac{1}{\mathsf{N}(\mathfrak{b})^s} \sum_{0 \neq a \in \mathfrak{b}^{-1}/R^\times} \frac{1}{\operatorname{Nm}(a)^s}. \tag{26.2.10}$$

One now reduces to a problem concerning lattice points in a fundamental domain for the action of $R^\times$, and examining the residue of the pole at $s = 1$ fits into a more general framework (invoked again below).

Definition 26.2.11. A **cone** $X \subseteq \mathbb{R}^n$ is a subset closed under multiplication by positive scalars, so $tX = X$ for all $t \in \mathbb{R}_{>0}$.

Theorem 26.2.12. *Let $X \subseteq \mathbb{R}^n$ be a cone. Let $N : X \to \mathbb{R}_{>0}$ be a function satisfying*

$$N(tx) = t^n N(x) \quad \text{for all } x \in X, t \in \mathbb{R}_{>0}.$$

Suppose that

$$X_{\le 1} := \{x \in X : N(x) \le 1\} \subseteq \mathbb{R}^n \tag{26.2.13}$$

is a bounded subset with volume $\operatorname{vol}(X_{\le 1})$. *Let $\Lambda \subseteq \mathbb{R}^n$ be a (full) $\mathbb{Z}$-lattice in $\mathbb{R}^n$, and let*

$$\zeta_{\Lambda,X}(s) := \sum_{\lambda \in X \cap \Lambda} \frac{1}{N(\lambda)^s}.$$

Then $\zeta_{\Lambda,X}(s)$ converges for $\operatorname{Re} s > 1$ *and has a simple pole at $s = 1$ with residue*

$$\zeta^*_{\Lambda,X}(1) = \lim_{s \searrow 1}(s-1)\zeta_{\Lambda,X}(s) = \frac{\operatorname{vol}(X_{\le 1})}{\operatorname{covol}(\Lambda)}.$$

Proof. See Borevich–Shafarevich [BS66, Chapter 5, Section 1.1, Theorem 1]. □

26.2.14. To apply Theorem 26.2.12 for $\zeta_{F,[\mathfrak{b}]}(s)$, we embed $F \hookrightarrow F_{\mathbb{R}} \simeq \mathbb{R}^r \times \mathbb{C}^c$ and we equip $F_{\mathbb{R}}$ with the inner product

$$\langle x, y \rangle = \sum_{i=1}^{r} x_i y_i + \sum_{j=1}^{c} 2\operatorname{Re}(x_{r+j}\overline{y_{r+j}}) \tag{26.2.15}$$

for $x = (x_i)_i, y = (y_i)_i \in F_{\mathbb{R}}$. This inner product modifies the usual one by rescaling complex coordinates, and the volume form vol induced by $\langle\,,\,\rangle$ is 2^c times the standard Lebesgue volume on $\mathbb{R}^r \times \mathbb{C}^c$. With this convention, we have $\langle x, 1\rangle = \operatorname{Tr}_{F|\mathbb{Q}}(x)$ and $\operatorname{covol}(R) = \sqrt{|d_F|}$.

We then take Λ to be the image of $\mathfrak{b}^{-1}$, and take X to be a cone fundamental domain for the action of the unit group $R^\times$. The absolute norm $\mathsf{N}(x) = |\operatorname{Nm}_{F|\mathbb{Q}}(x)|$ then satisfies the required homogeneity property, and $X_{\le 1}$ is bounded, so by Theorem 26.2.12,

$$\zeta^*_{F,[\mathfrak{b}]}(1) = \frac{1}{\mathsf{N}(\mathfrak{b})}\frac{\operatorname{vol}(X_{\le 1})}{\operatorname{covol}(\Lambda)}. \tag{26.2.16}$$

We have

$$\operatorname{covol}(\Lambda) = \frac{\operatorname{covol}(R)}{\mathsf{N}(\mathfrak{b})} = \frac{\sqrt{|d_F|}}{\mathsf{N}(\mathfrak{b})}. \tag{26.2.17}$$

It requires a bit more work to compute $\operatorname{vol}(X_{\le 1})$.

Proposition 26.2.18. *We have*

$$\operatorname{vol}(X_{\le 1}) = \frac{2^r(2\pi)^c \operatorname{Reg}_F}{w_F} \tag{26.2.19}$$

Proof. See Exercise 26.3: the proof is well-summarized as a "change of variables", but the reader may prefer the idelic point of view (Chapter 29) instead, where the integrals are 'easier'. A detailed proof can be found in Borevich–Shafarevich [BS66, §5.1.3], Lang [Lang94, §VI.3, Theorem 3], and Neukirch [Neu99, §VII.5]. □

Plugging (26.2.19) and (26.2.17) into (26.2.16),

$$\zeta^*_{F,[\mathfrak{b}]}(1) = \frac{2^r(2\pi)^c}{w_F\sqrt{|d_F|}}\operatorname{Reg}_F; \tag{26.2.20}$$

note in particular that this does not depend on the class $[\mathfrak{b}]$! The analytic class number formula (Theorem 26.2.3) then follows as

$$\zeta^*_F(1) = \sum_{[\mathfrak{b}]} \zeta^*_{F,[\mathfrak{b}]}(1) = \frac{2^r(2\pi)^c}{w_F\sqrt{|d_F|}}\operatorname{Reg}_F h_F. \tag{26.2.21}$$

26.3 Classical zeta functions of quaternion algebras

We now embark on a proof in our quaternionic setting, mimicking the above. We retain our notation on the number field F. We further let throughout B be a quaternion algebra over F of discriminant $\mathfrak{D}$ and let $O \subseteq B$ be an R-order. (Our emphasis will be on the case O a maximal order, but many definitions carry through.)

To begin, in this section we define the classical zeta function and show it has an Euler product.

26.3.1. Let I be an invertible, integral right O-ideal, so that $I \subseteq O$, and by definition I is sated so $O_{\mathsf{R}}(I) = O$. Recall we have defined $\mathsf{N}(I) = \#(O/I)$; we have $\mathsf{N}(I) = \mathsf{N}(\operatorname{nrd}(I))^2$ (Paragraph 16.4.10).

For example, if $\mathfrak{a} \subseteq R$ is a nonzero ideal then $\mathsf{N}(\mathfrak{a}O) = \mathsf{N}(\mathfrak{a})^4$.

We then define the **(classical) zeta function** of O to be

$$\zeta_O(s) := \sum_{I \subseteq O} \frac{1}{\mathsf{N}(I)^s} = \sum_{\mathfrak{n}} \frac{a_{\mathfrak{n}}(O)}{\mathsf{N}(\mathfrak{n})^{2s}} \tag{26.3.2}$$

where the first sum is over all (nonzero) integral, invertible right O-ideals I and in the second sum we define

$$a_{\mathfrak{n}}(O) := \#\{I \subseteq O : \operatorname{nrd}(I) = \mathfrak{n}\} \tag{26.3.3}$$

(and $a_{\mathfrak{n}}(O)$ is finite by Lemma 17.7.26).

Lemma 26.3.4. *If O, O' are locally isomorphic, then $a_{\mathfrak{n}}(O) = a_{\mathfrak{n}}(O')$ for all $\mathfrak{n}$.*

Proof. We use the local-global dictionary for lattices (Theorem 9.4.9). To ease parentheses in the notation, we work in the completion, but one can also work just in the localization. For all $\mathfrak{p}$, we have $O'_{\mathfrak{p}} = \nu_{\mathfrak{p}}^{-1} O_{\mathfrak{p}} \nu_{\mathfrak{p}}$ for some $\nu_{\mathfrak{p}} \in B_{\mathfrak{p}}^{\times}$, and we may take $\nu_{\mathfrak{p}} = 1$ for all but finitely many $\mathfrak{p}$; the element $\nu_{\mathfrak{p}}$ is well-defined up to left multiplication by $O_{\mathfrak{p}}^{\times}$ and right multiplication by $O_{\mathfrak{p}}'^{\times}$.

Then to an integral, invertible right O-ideal I, we associate the unique lattice I' such that $I'_{\mathfrak{p}} = \nu_{\mathfrak{p}}^{-1} I_{\mathfrak{p}} \nu_{\mathfrak{p}}$; such a lattice is well-defined independent of the choice of $\nu_{\mathfrak{p}}$. By construction, $O_{\mathsf{R}}(I'_{\mathfrak{p}}) = O'_{\mathfrak{p}}$ so $O_{\mathsf{R}}(I') = O'$. And since I is integral, $I_{\mathfrak{p}} \subseteq O_{\mathfrak{p}}$ whence $I'_{\mathfrak{p}} \subseteq \nu_{\mathfrak{p}}^{-1} I_{\mathfrak{p}} \nu_{\mathfrak{p}} \subseteq O'_{\mathfrak{p}}$ and I is locally integral hence integral. Since I is invertible, I is locally principal, so I' is also locally principal, hence invertible. Finally, again checking locally, we have $\operatorname{nrd}(I') = \operatorname{nrd}(I)$.

Repeating this argument going from I' to I, we see that the corresponding sets of ideals are in bijection, as claimed. □

26.3.5. From Lemma 26.3.4, we see that $\zeta_O(s)$ only depends on the genus of O. Since there is a unique genus of maximal orders in B, following the number field case we will write $\zeta_B(s) = \zeta_O(s)$ where O is any maximal order.

Our next order of business is to establish an Euler product for $\zeta_O(s)$. We prove a more general result on the factorization of invertible lattices.

Lemma 26.3.6. *Let I be an invertible, integral lattice and suppose that* $\operatorname{nrd}(I) = \mathfrak{m}\mathfrak{n}$ *with $\mathfrak{m}, \mathfrak{n} \subseteq R$ coprime ideals. Then there exists a unique invertible, integral lattice J such that I is compatible with J^{-1} with IJ^{-1} integral and* $\operatorname{nrd}(J) = \mathfrak{m}$.

Proof. We use the local-global dictionary for lattices, and we define $J \subseteq B$ to be the unique lattice such that

$$J_{(\mathfrak{p})} := \begin{cases} I_{(\mathfrak{p})} = O_{(\mathfrak{p})}, & \text{if } \mathfrak{p} \nmid \mathfrak{m}\mathfrak{n}; \\ I_{(\mathfrak{p})}, & \text{if } \mathfrak{p} \mid \mathfrak{m}; \\ O_{(\mathfrak{p})}, & \text{if } \mathfrak{p} \mid \mathfrak{n}. \end{cases} \tag{26.3.7}$$

We have $O_{\mathsf{R}}(J) = O$ and $\operatorname{nrd}(J) = \mathfrak{m}$, since these statements hold locally. Integrality and invertibility are local; since these are true for I they are true for J. Finally, we compute that $(IJ^{-1})_{(\mathfrak{p})} = O_{(\mathfrak{p})}$ for all $\mathfrak{p} \nmid \mathfrak{n}$ and $(IJ^{-1})_{(\mathfrak{p})} = I_{(\mathfrak{p})}$ for $\mathfrak{p} \mid \mathfrak{n}$, so IJ^{-1} is locally integral and hence integral. The uniqueness of J can be verified directly (Exercise 26.4). □

26.3.8. Consider the situation of Lemma 26.3.6. Let $I' = IJ^{-1}$. Then $I = I'J$, and I' is integral, invertible (Paragraph 16.5.3) and compatible with J. Since $\operatorname{nrd}(I') = \mathfrak{n}$, we have "factored" I.

We have $O_{\mathsf{R}}(I') = O_{\mathsf{L}}(J)$ by compatibility, but this common order is only locally isomorphic to O, since I', J are locally principal but not necessarily principal. So in a sense, this factorization occurs not over O but over the genus of O—but this is a harmless extension.

Proposition 26.3.9. *If $\mathfrak{m}, \mathfrak{n}$ are coprime, then $a_{\mathfrak{m}\mathfrak{n}}(O) = a_{\mathfrak{m}}(O)a_{\mathfrak{n}}(O)$.*

Proof. Write $A_{\mathfrak{n}}(O)$ for the set of integral, invertible right O-ideals I with $\operatorname{nrd}(I) = \mathfrak{n}$. Then $\#A_{\mathfrak{n}}(O) = a_{\mathfrak{n}}(O)$. According to Lemma 26.3.6, there is a map

$$\begin{aligned} A_{\mathfrak{m}\mathfrak{n}}(O) &\to A_{\mathfrak{n}}(O) \\ I &\mapsto J \end{aligned} \tag{26.3.10}$$

We claim that this map is surjective and that each fiber has cardinality $a_{\mathfrak{m}}(O)$. Indeed, these statements follow at the same time from the following observation: if $J \in A_{\mathfrak{n}}(O)$ with $O' = O_{\mathsf{L}}(J)$, then for each $I' \in A_{\mathfrak{m}}(O')$, we have $\overline{I'}$ compatible with J and $I = I'J \in A_{\mathfrak{m}}$, and conversely; so the fiber of (26.3.10) is identified with $A_{\mathfrak{m}}(O')$, of cardinality $a_{\mathfrak{m}}(O') = a_{\mathfrak{m}}(O)$ by Lemma 26.3.4. □

26.3.11. From Proposition 26.3.9 and unique factorization of ideals in R, we find that ζ_O has an **Euler product**

$$\zeta_O(s) = \prod_{\mathfrak{p}} \zeta_{O_{\mathfrak{p}}}(s) \tag{26.3.12}$$

where

$$\zeta_{O_{\mathfrak{p}}}(s) := \sum_{I_{\mathfrak{p}} \subseteq O_{\mathfrak{p}}} \frac{1}{\mathsf{N}(I_{\mathfrak{p}})^s} = \sum_{e=0}^{\infty} \frac{a_{\mathfrak{p}^e}(O)}{\mathsf{N}(\mathfrak{p})^{2s}}. \tag{26.3.13}$$

Remark 26.3.14. Zeta functions of semisimple algebras over a number field can be defined in the same way as in (26.3.2), following Solomon [Sol77]: see the survey on analytic methods in noncommutative number theory by Bushnell–Reiner [BR85].

Remark 26.3.15. The world of L-functions is rich and very deep: for a beautiful survey of the analytic theory of automorphic L-functions in historical perspective, see Gelbart–Miller [GM2004]. In particular, we have not given a general definition of zeta functions (or L-functions) in this section, but it is generally agreed that the *Selberg class* incorporates the minimal essential features: definition as a Dirichlet series, meromorphic continuation to the complex plane, Euler product, and functional equation. See e.g. Conrey–Ghosh [CG93] and the references therein.

26.4 Counting ideals in a maximal order

We now count ideals of prime power norm. By the local-global dictionary, there is a bijection

$$\{I \subseteq O : \operatorname{nrd}(I) = \mathfrak{p}^e\} \xrightarrow{\sim} \{I_{\mathfrak{p}} \subseteq O_{\mathfrak{p}} : \operatorname{nrd}(I_{\mathfrak{p}}) = \mathfrak{p}^e\}.$$

so it suffices to count the number of ideals in the local case. In this section, we carry out this count for maximal orders.

So let $\mathfrak{p} \subset R$ be a (nonzero) prime and let $q := \operatorname{Nm}(\mathfrak{p})$. Let $O_{\mathfrak{p}} \subset B_{\mathfrak{p}}$ be a maximal order. Let

$$a_{\mathfrak{p}^e}(O_{\mathfrak{p}}) = \#\{I_{\mathfrak{p}} = \alpha_{\mathfrak{p}} O_{\mathfrak{p}} \subseteq O_{\mathfrak{p}} : \operatorname{nrd}(I_{\mathfrak{p}}) = \mathfrak{p}^e\}$$

count the number of right integral $O_{\mathfrak{p}}$-ideals of norm $\mathfrak{p}^e$. Since $O_{\mathfrak{p}}$ is maximal, every nonzero ideal is invertible; and because $R_{\mathfrak{p}}$ is a DVR, all such invertible ideals are principal.

Lemma 26.4.1. *Let $O_{\mathfrak{p}} \subset B_{\mathfrak{p}}$ be a maximal order and let $e \in \mathbb{Z}_{\geq 0}$.*

(a) *If $B_{\mathfrak{p}}$ is a division ring, then every right integral $O_{\mathfrak{p}}$-ideal is a power of the maximal ideal and $a_{\mathfrak{p}^e}(O_{\mathfrak{p}}) = 1$.*

(b) *If $B_{\mathfrak{p}} \simeq \mathrm{M}_2(F_{\mathfrak{p}})$, so that $O_{\mathfrak{p}} \simeq \mathrm{M}_2(R_{\mathfrak{p}})$, then the set of right integral $O_{\mathfrak{p}}$-ideals of reduced norm $\mathfrak{p}^e$ is in bijection with the set*

$$\left\{ \begin{pmatrix} \pi^u & 0 \\ c & \pi^v \end{pmatrix} : u, v \in \mathbb{Z}_{\geq 0}, u + v = e \text{ and } c \in R/\mathfrak{p}^v \right\}$$

and

$$a_{\mathfrak{p}^e}(O_{\mathfrak{p}}) = 1 + q + \cdots + q^e. \tag{26.4.2}$$

Proof. For (a), if $\mathfrak{p}$ is ramified then by the work of section 13.3, there is a unique maximal order $O_{\mathfrak{p}}$ with a unique (two-sided) maximal ideal $J_{\mathfrak{p}}$ having $\mathrm{nrd}(J_{\mathfrak{p}}) = \mathfrak{p}$, and all ideals of $O_{\mathfrak{p}}$ are powers of $J_{\mathfrak{p}}$.

To prove (b), we appeal to the theory of elementary divisors (applying column operations, acting on the right). Suppose $O_{\mathfrak{p}} = \mathrm{M}_2(R_{\mathfrak{p}})$. Let $I_{\mathfrak{p}} = \alpha_{\mathfrak{p}} O_{\mathfrak{p}}$ be a right integral $O_{\mathfrak{p}}$-ideal of norm $\mathfrak{p}^e$ and let π be a uniformizer for $\mathfrak{p}$. Then by the theory of elementary divisors, we can write

$$\alpha_{\mathfrak{p}} = \begin{pmatrix} \pi^u & 0 \\ c & \pi^v \end{pmatrix}$$

for unique $u, v \in \mathbb{Z}_{\geq 0}$ with $u + v = e$ and $c \in R$ is uniquely defined as element of $R/\mathfrak{p}^v$ (Exercise 26.6). It follows that the number of such ideals is equal to $\sum_{v=0}^{e} q^v = 1 + q + \cdots + q^e$. ∎

26.4.3. There is an alternate bijection that is quite useful. We say an integral right O-ideal I is $\mathfrak{p}$-**primitive** if it does not contain $\mathfrak{p}O$ (so we cannot write $I = \mathfrak{p}I'$ with I' integral).

For a commutative ring A, we define the **projective line** over A to be the set

$$\mathbb{P}^1(A) := \{(x, y) \in A^2 : xA + yA = A\}/A^\times$$

and write equivalence classes $(x : y) \in \mathbb{P}^1(A)$.

Then for $O_{\mathfrak{p}} = \mathrm{M}_2(R_{\mathfrak{p}})$, there is a bijection

$$\begin{aligned} \mathbb{P}^1(R/\mathfrak{p}^e) &\to \{I_{\mathfrak{p}} \subseteq O_{\mathfrak{p}} : I_{\mathfrak{p}} \text{ primitive and } \mathrm{nrd}(I_{\mathfrak{p}}) = \mathfrak{p}^e\} \\ (a : c) &\mapsto \begin{pmatrix} a & 0 \\ c & 0 \end{pmatrix} O_{\mathfrak{p}} + \mathfrak{p}^e O_{\mathfrak{p}} \end{aligned} \tag{26.4.4}$$

Any ideal of the form in the right-hand side of (26.4.4) is a primitive right integral $O_{\mathfrak{p}}$-ideal with reduced norm $\mathfrak{p}^e$. Conversely, suppose that $I_{\mathfrak{p}} = \alpha_{\mathfrak{p}} O_{\mathfrak{p}}$ is primitive. We have $\mathrm{nrd}(\alpha_{\mathfrak{p}}) \equiv 0 \pmod{\mathfrak{p}^e}$. We find a "standard form" for $I_{\mathfrak{p}}$ by looking at the left kernel of $\alpha_{\mathfrak{p}}$. Let

$$L := \{x \in (R/\mathfrak{p}^e)^2 : x\alpha_{\mathfrak{p}} \equiv 0 \pmod{\mathfrak{p}^e}\}.$$

We claim that L is a free $R/\mathfrak{p}^e$-module of rank 1. Indeed, L is one-dimensional over $R/\mathfrak{p}$ since $I_{\mathfrak{p}}$ is primitive and so $\alpha_{\mathfrak{p}} \not\equiv 0 \pmod{\mathfrak{p}}$; by Hensel's lemma, it follows that L is also one-dimensional. Therefore, there is a unique generator $(a : c) \in \mathbb{P}^1(R/\mathfrak{p}^e)$ for L. We therefore define an map $I_{\mathfrak{p}} \mapsto (-c : a)$ and verify that this furnishes an inverse to (26.4.4).

Since $\#\mathbb{P}^1(R/\mathfrak{p}^e) = q^e + q^{e-1}$ for $e \geq 1$, we recover the count (26.4.2) as

$$a_{\mathfrak{p}^e}(O_\mathfrak{p}) = \sum_{i=0}^{\lfloor e/2 \rfloor} \#\mathbb{P}^1(R/\mathfrak{p}^{e-2i}) = q^e + q^{e-1} + \cdots + q + 1.$$

26.4.5. Lemma 26.4.1 implies a factorization of $\zeta_{B_p}(s) = \zeta_{O_\mathfrak{p}}(s)$. Write

$$\zeta_{F_p}(s) = \sum_{e=0}^{\infty} \frac{1}{q^{es}} = \left(1 - \frac{1}{q^s}\right)^{-1} \tag{26.4.6}$$

so that $\zeta_F(s) = \prod_\mathfrak{p} \zeta_{F_p}(s)$.

Corollary 26.4.7. *We have*

$$\zeta_{B_\mathfrak{p}}(s) = \left(1 - \frac{1}{q^{2s}}\right)^{-1} \cdot \begin{cases} 1, & \textit{if } \mathfrak{p} \textit{ is ramified;} \\ \left(1 - 1/q^{2s-1}\right)^{-1}, & \textit{if } \mathfrak{p} \textit{ is split.} \end{cases}$$

Equivalently,

$$\zeta_{B_\mathfrak{p}}(s) = \begin{cases} \zeta_{F_\mathfrak{p}}(2s), & \textit{if } \mathfrak{p} \textit{ is ramified;} \\ \zeta_{F_\mathfrak{p}}(2s)\zeta_{F_\mathfrak{p}}(2s-1), & \textit{if } \mathfrak{p} \textit{ is split.} \end{cases}$$

Proof. We use Lemma 26.4.1. If $B_\mathfrak{p}$ is a division ring, then Lemma 26.4.1(a) applies, and the result is immediate. For the second case, we compute

$$\begin{aligned} \zeta_{B_\mathfrak{p}}(s) &= \sum_{e=0}^{\infty} \frac{1 + q + \cdots + q^e}{q^{2es}} = \sum_{e=0}^{\infty} \frac{1 - q^{e+1}}{(1-q)q^{2es}} \\ &= \frac{1}{1-q}\left(\sum_{e=0}^{\infty} \frac{1}{q^{2es}} - q \sum_{e=0}^{\infty} \frac{1}{q^{(2s-1)e}}\right) \\ &= \frac{1}{1-q}\left(\frac{1}{1 - 1/q^{2s}} - \frac{q}{1 - 1/q^{2s-1}}\right) \\ &= \left(1 - \frac{1}{q^{2s}}\right)^{-1}\left(1 - \frac{1}{q^{2s-1}}\right)^{-1} \end{aligned} \tag{26.4.8}$$

as claimed. ∎

We have proven the following result.

Theorem 26.4.9 (Factorization of $\zeta_B(s)$, maximal order). *Let B be a quaternion algebra of discriminant $\mathfrak{D} = \operatorname{disc} B$. Then*

$$\zeta_B(s) = \prod_\mathfrak{p} \zeta_{B_\mathfrak{p}}(s) = \zeta_F(2s)\zeta_F(2s-1) \prod_{\mathfrak{p} \mid \mathfrak{D}} (1 - \mathsf{N}(\mathfrak{p})^{1-2s}). \tag{26.4.10}$$

Proof. Combine the Euler product 26.3.11 with Corollary 26.4.7. ∎

Corollary 26.4.11. *$\zeta_B(s)$ has a simple pole at $s = 1$ with residue*

$$\zeta_B^*(1) = \lim_{s\to 1}(s-1)\zeta_B(s) = \zeta_F(2)\frac{\zeta_F^*(1)}{2}\prod_{\mathfrak{p}\mid\mathfrak{D}}(1-\mathsf{N}(\mathfrak{p})^{-1}). \tag{26.4.12}$$

Proof. Since $\zeta_F(s)$ has only a simple pole at $s = 1$, with residue computed in Theorem 26.2.3, there is a single simple pole of $\zeta_B(s)$ at $s = 1$. □

26.5 Eichler mass formula: maximal orders

We now finish the proof of the Eichler mass formula (Main Theorem 26.1.5) for maximal orders (26.1.13). In the next section, we will deduce the general formula from it: for a nonmaximal order, there are extra factors at each prime dividing the discriminant, and it is simpler to account for those in a separate step.

In this section, we now suppose that B is *definite*, so F is a totally real field (and $d_F > 0$). In particular, B is a division algebra. We saw in 26.1.3 that it was natural to weight ideal classes inversely by the size of their automorphism group (modulo scalars). To this end, and noting $R^\times \trianglelefteq \mathcal{O}^\times$ is central so normal, we prove the following lemma.

Lemma 26.5.1. *The group $\mathcal{O}^\times/R^\times$ is finite.*

Proof. In Lemma 17.7.13, we proved that

$$\mathcal{O}^1 := \{\gamma \in \mathcal{O}^\times : \operatorname{nrd}(\gamma) = 1\}$$

is a finite group by embedding $\mathcal{O} \hookrightarrow B_\mathbb{R} \simeq \mathbb{R}^{4n}$ as a Euclidean lattice with respect to the absolute reduced norm (see 17.7.10). Since $\mathcal{O}^1 \cap R^\times = \{\pm 1\}$, the reduced norm gives an exact sequence

$$1 \to \frac{\mathcal{O}^1}{\{\pm 1\}} \to \frac{\mathcal{O}^\times}{R^\times} \xrightarrow{\operatorname{nrd}} \frac{R^\times}{R^{\times 2}}. \tag{26.5.2}$$

By Dirichlet's unit theorem, the group $R^\times$ is finitely generated (of rank $r + c - 1$), so the group $R^\times/R^{\times 2}$ is a finite abelian 2-group. The result follows. □

We will examine unit groups in detail in Chapter 32. With this finiteness statement in hand, we make the following definition.

Definition 26.5.3. Define the **mass** of $\mathcal{O}$ to be

$$\operatorname{mass}(\operatorname{Cls}\mathcal{O}) := \sum_{[J]\in\operatorname{Cls}\mathcal{O}} \frac{1}{w_J}$$

where $w_J = [\mathcal{O}_\mathsf{L}(J)^\times : R^\times] \in \mathbb{Z}_{\geq 1}$.

Theorem 26.5.4 (Eichler's mass formula). *Let O be a maximal order in a totally definite quaternion algebra B of discriminant $\mathfrak{D}$. Then*

$$\operatorname{mass}(\operatorname{Cls} O) = \frac{2}{(2\pi)^{2n}} h_F d_F^{3/2} \varphi(\mathfrak{D})$$

where $\varphi(\mathfrak{D}) = \prod_{\mathfrak{p}\mid\mathfrak{D}}(\mathsf{N}(\mathfrak{p}) - 1)$.

Following the strategy in the classical case (to prove the analytic class number formula), to prove Theorem 26.5.4 we will write $\zeta_O(s)$ as a sum over right ideal classes and analyze its residue at $s = 1$ by a volume computation.

26.5.5. For an integral invertible right O-ideal J, let

$$\zeta_{O,[J]}(s) := \sum_{\substack{I\subseteq O\\ [I]=[J]}} \frac{1}{\mathsf{N}(I)^s}. \tag{26.5.6}$$

Then

$$\zeta_O(s) = \sum_{[J]\in\operatorname{Cls} O} \zeta_{O,[J]}(s).$$

We have $[I] = [J]$ if and only if $I \simeq J$ if and only if $I = \alpha J$ for nonzero $\alpha \in J^{-1}$. Since $\mu J = J$ if and only if $\mu \in O_{\mathsf{L}}(J)^\times$ (Exercise 16.3), it follows that

$$\zeta_{O,[J]}(s) = \frac{1}{\mathsf{N}(J)^s} \sum_{0\neq\alpha\in J^{-1}/O_{\mathsf{L}}(J)^\times} \frac{1}{\mathsf{N}(\alpha)^s}. \tag{26.5.7}$$

By Lemma 26.5.1, we have

$$w_J := [O_{\mathsf{L}}(J)^\times : R^\times] \in \mathbb{Z}_{>0}. \tag{26.5.8}$$

Then (26.5.7) becomes

$$\zeta_{O,[J]}(s) = \frac{1}{w_J \mathsf{N}(J)^s} \sum_{0\neq\alpha\in J^{-1}/R^\times} \frac{1}{\mathsf{N}(\alpha)^s}. \tag{26.5.9}$$

Proposition 26.5.10. *Let* $\mathfrak{N} := \operatorname{discrd}(O)$. *Then* $\zeta_{O,[J]}(s)$ *has a simple pole at* $s = 1$ *with residue*

$$\zeta^*_{O,[J]}(1) = \frac{2^n(2\pi)^{2n}\operatorname{Reg}_F}{8w_J d_F^2 \mathsf{N}(\mathfrak{N})}.$$

Proof. We relate residue to volumes using Theorem 26.2.12. We recall (again) 17.7.10: this gives

$$J^{-1} \hookrightarrow B \hookrightarrow B_{\mathbb{R}} := B \otimes_{\mathbb{Q}} \mathbb{R} \simeq \mathbb{H}^n \simeq \mathbb{R}^{4n}$$

the structure of a Euclidean lattice $\Lambda \subseteq \mathbb{R}^{4n}$ with respect to the absolute reduced norm. We take the function N in Theorem 26.2.12 to be the absolute norm N (recalling 16.4.8).

We claim that

$$\operatorname{covol}(O) = \frac{d_F^2 \mathsf{N}(\mathfrak{N})}{2^n}. \tag{26.5.11}$$

By compatible real scaling, it is enough to prove that this relation holds for a single order O, and we choose the R-order

$$O = R \oplus Ri \oplus Rj \oplus Rk. \tag{26.5.12}$$

The lattice $R \subseteq F_{\mathbb{R}}$ has covolume $\sqrt{d_F}$, so R^4 has covolume $\sqrt{d_F}^4 = d_F^2$; the $\mathbb{Z}$-order $\mathbb{Z}\oplus\mathbb{Z}i\oplus\mathbb{Z}j\oplus\mathbb{Z}k$ has reduced discriminant 4 and covolume 1; and putting these together, the formula (26.5.11) is verified.

Then (26.5.11) and $\mathsf{N}(J) = [O : J]_{\mathbb{Z}} = [J^{-1} : O]$ imply that

$$\operatorname{covol}(\Lambda) = \frac{\operatorname{covol}(O)}{\mathsf{N}(J)} = \frac{d_F^2 \mathsf{N}(\mathfrak{N})}{2^n \mathsf{N}(J)}. \tag{26.5.13}$$

Next, the group $O_{\mathsf{L}}(J)^\times$ acts on J^{-1} (and on $B_{\mathbb{R}}$); and this group contains $R^\times$ with finite index $w_J = [O_{\mathsf{L}}(J) : R^\times]$, so

$$\operatorname{vol}(O_{\mathsf{L}}(J)^\times \backslash B_{\mathbb{R}}) = \frac{1}{w_J} \operatorname{vol}(R^\times \backslash B_{\mathbb{R}}). \tag{26.5.14}$$

Multiplication provides an identification

$$B_{\mathbb{R},\leq 1} \simeq F_{\mathbb{R},\leq 1} \times (\mathbb{H}^1)^n,$$

so

$$X_{\leq 1} = R^\times \backslash B_{\mathbb{R},\leq 1} \simeq (E \backslash F_{\mathbb{R},\leq 1}) \times (\{\pm 1\} \backslash (\mathbb{H}^1)^n) \tag{26.5.15}$$

where $E \leq R^\times$ is acting by squares. Thus

$$\operatorname{vol}(R^\times \backslash F_{\mathbb{R},\leq 1}) = \frac{2^{n-1}}{2(2^n)} \operatorname{Reg}_F = \frac{1}{4} \operatorname{Reg}_F. \tag{26.5.16}$$

Therefore

$$\operatorname{vol}(X_{\leq 1}) = \frac{(2\pi^2)^n \operatorname{Reg}_F}{8 w_J}. \tag{26.5.17}$$

From Theorem 26.2.12 together with (26.5.13) and (26.5.17),

$$\zeta^*_{O,[J]}(1) = \frac{4^n (2\pi^2)^n \operatorname{Reg}_F}{8 w_J d_F^2 \mathsf{N}(\mathfrak{N})} = \frac{2^n (2\pi)^{2n} \operatorname{Reg}_F}{8 w_J d_F^2 \mathsf{N}(\mathfrak{N})}. \tag{26.5.18}$$

□

We now conclude the proof.

Proof of Theorem 26.5.4. We now suppose that $O \subset B$ is a maximal order, and write $\zeta_B(s)$ and $\zeta_{B,[J]}(s)$. We compare the evaluation of residues given by Corollary 26.4.11 and Proposition 26.5.10. Since $\zeta_B(s)$ and each $\zeta_{B,[J]}(s)$ have simple poles at $s = 1$, we get

$$\zeta^*_B(1) = \sum_{[J] \in \operatorname{Cls} O} \zeta^*_{B,[J]}(1).$$

From (26.4.12),

$$\zeta_B^*(1) = \zeta_F(2)\frac{\zeta_F^*(1)}{2}\prod_{\mathfrak{p}\mid\mathfrak{D}}\left(1-\frac{1}{\mathsf{N}(\mathfrak{p})}\right) = \zeta_F(2)\frac{\zeta_F^*(1)}{2}\frac{\varphi(\mathfrak{D})}{\mathsf{N}(\mathfrak{D})}. \tag{26.5.19}$$

From the analytic class number formula (Theorem 26.2.3),

$$\zeta_F^*(1) = \frac{2^n}{2\sqrt{d_F}}h_F\operatorname{Reg}_F$$

since $w_F = 2$ (as F is totally real).

Adding the residues from Lemma 26.5.10, we find that

$$\frac{2^n\zeta_F(2)}{4\sqrt{d_F}}h_F\operatorname{Reg}_F\frac{\varphi(\mathfrak{D})}{\mathsf{N}(\mathfrak{D})} = \frac{2^n(2\pi)^{2n}\operatorname{Reg}_F}{8d_F^2\mathsf{N}(\mathfrak{D})}\sum_{[J]\in\operatorname{Cls}\mathcal{O}}\frac{1}{w_J}. \tag{26.5.20}$$

Cancelling, we find

$$\operatorname{mass}(\operatorname{Cls}\mathcal{O}) = \sum_{[J]\in\operatorname{Cls}\mathcal{O}}\frac{1}{w_J} = \frac{2}{(2\pi)^{2n}}\zeta_F(2)d_F^{3/2}h_F\varphi(\mathfrak{D}) \tag{26.5.21}$$

and this concludes the proof. □

Remark 26.5.22. For an alternative direct approach in this setting using Epstein zeta functions, see Sands [San2017].

26.6 Eichler mass formula: general case

We now consider the general case of the Eichler mass formula, involving two steps. First, we relate the class set of a suborder to the class set of a (maximal) superorder; second, we compute the fibers of this map via a group action of the units.

For these steps, we refresh our notation and allow B to be a definite or indefinite quaternion algebra over F.

26.6.1. Let $\mathcal{O}' \supseteq \mathcal{O}$ be an R-superorder, and suppose that there is a prime $\mathfrak{p}$ such that $\mathcal{O}'_{\mathfrak{q}} = \mathcal{O}_{\mathfrak{q}}$ for all primes $\mathfrak{q} \neq \mathfrak{p}$. We refine the map from Exercise 17.3(b) as follows. For $I \subseteq \mathcal{O}$ a right $\mathcal{O}$-ideal, we define the right $\mathcal{O}'$-ideal $\rho(I) = I\mathcal{O}' \subseteq \mathcal{O}'$ obtained by extension. Then ρ induces a map

$$\begin{aligned}\operatorname{Cls}\mathcal{O} &\to \operatorname{Cls}\mathcal{O}'\\ [I] &\mapsto [I\mathcal{O}']\end{aligned} \tag{26.6.2}$$

that is well-defined and surjective (Exercise 26.5(a)). Let $[I'] \in \operatorname{Cls}\mathcal{O}'$ and consider the set

$$\rho^{-1}(I') = \{I \subseteq \mathcal{O} : I\mathcal{O} = I'\},$$

the fiber of the extension map over I'.

We define an action of the group $O_{\mathfrak{p}}'^{\times}$ on $\rho^{-1}(I')$ as follows. Write $I'_{\mathfrak{p}} = \beta_{\mathfrak{p}} O'_{\mathfrak{p}}$. Then to $\mu_{\mathfrak{p}} \in O_{\mathfrak{p}}'^{\times}$, we associate the unique lattice $I\langle \mu_{\mathfrak{p}} \rangle$ (the notation to suggest "the lattice generated by $\mu_{\mathfrak{p}}$") such that

$$I\langle \mu_{\mathfrak{p}} \rangle_{\mathfrak{p}} = \beta_{\mathfrak{p}} \mu_{\mathfrak{p}} O_{\mathfrak{p}}$$

and $I\langle \mu_{\mathfrak{p}} \rangle_{\mathfrak{q}} = I_{\mathfrak{q}} = I'_{\mathfrak{q}}$ for all $\mathfrak{q} \neq \mathfrak{p}$, using the local-global dictionary (Theorem 9.4.9). This defines a right action of $O_{\mathfrak{p}}'^{\times}$; it acts simply transitively on $\rho^{-1}(I')$, and the kernel of this action is visibly the subgroup $O_{\mathfrak{p}}^{\times}$. Therefore

$$\#\rho^{-1}(I') = [O_{\mathfrak{p}}'^{\times} : O_{\mathfrak{p}}^{\times}].$$

We now look at the classes in the fiber. If $\mu_{\mathfrak{p}}, \nu_{\mathfrak{p}} \in O_{\mathfrak{p}}'^{\times}$ have $[I\langle \mu_{\mathfrak{p}} \rangle] = [I\langle \nu_{\mathfrak{p}} \rangle] \in \operatorname{Cls} O$, then there exists $\alpha \in B^{\times}$ such that

$$\alpha I\langle \mu_{\mathfrak{p}} \rangle = I\langle \nu_{\mathfrak{p}} \rangle$$

and by extension $\alpha I' = I'$, so $\alpha \in O_{\mathsf{L}}(I')$, and conversely. Therefore, we have a bijection

$$\operatorname{Cls} O \leftrightarrow \bigsqcup_{[I'] \in \operatorname{Cls} O'} O_{\mathsf{L}}(I')^{\times} \backslash \rho^{-1}(I'). \tag{26.6.3}$$

(See also Pacetti–Sirolli [PS2014, §3].)

Proposition 26.6.4. *Let $O' \supseteq O$ be an R-superorder, and suppose that there is a prime $\mathfrak{p}$ such that $O'_{\mathfrak{q}} = O_{\mathfrak{q}}$ for all primes $\mathfrak{q} \neq \mathfrak{p}$. Then*

$$\operatorname{mass}(\operatorname{Cls} O) = [O_{\mathfrak{p}}'^{\times} : O_{\mathfrak{p}}^{\times}] \operatorname{mass}(\operatorname{Cls} O').$$

Proof. By (26.6.3), we conclude that

$$\begin{aligned} \operatorname{mass}(\operatorname{Cls} O) &= \sum_{[I] \in \operatorname{Cls} O} \frac{1}{w_I} = \sum_{[I'] \in \operatorname{Cls} O'} \sum_{IO' = I'} \frac{1}{w_I} \left(\frac{w_{I'}}{w_I} \right)^{-1} \\ &= \sum_{[I'] \in \operatorname{Cls} O'} [O_{\mathfrak{p}}'^{\times} : O_{\mathfrak{p}}^{\times}] \frac{1}{w_{I'}} \\ &= [O_{\mathfrak{p}}'^{\times} : O_{\mathfrak{p}}^{\times}] \operatorname{mass}(\operatorname{Cls} O') \end{aligned} \tag{26.6.5}$$

as claimed. ∎

In order to apply Proposition 26.6.4, we need to compute the index of unit groups, a quantity that depends on the (locally defined) Eichler symbol. For a prime $\mathfrak{p}$, we define

$$\lambda(O, \mathfrak{p}) := \frac{1 - \operatorname{Nm}(\mathfrak{p})^{-2}}{1 - \left(\dfrac{O}{\mathfrak{p}} \right) \operatorname{Nm}(\mathfrak{p})^{-1}} = \begin{cases} 1 + 1/q, & \text{if } (O \mid p) = 1; \\ 1 - 1/q, & \text{if } (O \mid p) = -1; \\ 1 - 1/q^2, & \text{if } (O \mid p) = 0. \end{cases} \tag{26.6.6}$$

Lemma 26.6.7. *Let $O' \supseteq O$ be a containment of R-orders with O' maximal. Then*

$$[O'^{\times}_{\mathfrak{p}} : O^{\times}_{\mathfrak{p}}] = [O'_{\mathfrak{p}} : O_{\mathfrak{p}}]\lambda(O,\mathfrak{p}) \cdot \begin{cases} 1, & \text{if } \mathfrak{p} \text{ is split in } B; \\ (1-1/q)^{-1}, & \text{if } \mathfrak{p} \text{ is ramified in } B. \end{cases}$$

Proof. We follow Körner [Kör85, §3]. To prove the lemma, we may localize at $\mathfrak{p}$ and so we drop the subscripts. Let $n \in \mathbb{Z}_{\geq 1}$ be such that $\mathfrak{p}^n O' \subseteq \mathfrak{p}O$. Then

$$[O'^{\times} : O^{\times}] = \frac{[O'^{\times} : 1+\mathfrak{p}O'][1+\mathfrak{p}O' : 1+\mathfrak{p}^n O']}{[O^{\times} : 1+\mathfrak{p}O][1+\mathfrak{p}O : 1+\mathfrak{p}^n O']}.$$

For $\gamma, \delta \in 1+\mathfrak{p}O$, we have $\gamma\delta^{-1} \in 1+\mathfrak{p}^n O'$ if and only if $\gamma - \delta \in \mathfrak{p}^n O'$. Therefore

$$[1+\mathfrak{p}O : 1+\mathfrak{p}^n O'] = [\mathfrak{p}O : \mathfrak{p}^n O'] = [O : \mathfrak{p}^{n-1}O']$$

and similarly with O', all indices taken as abelian groups. Therefore

$$\frac{[1+\mathfrak{p}O' : 1+\mathfrak{p}^n O']}{[1+\mathfrak{p}O : 1+\mathfrak{p}^n O']} = [O' : O].$$

For the other terms, we recall Lemma 24.3.12. We divide up into the cases, noting that if$(O' \mid \mathfrak{p}) = -1$ then we must have $\varepsilon = -1, 0$ by classification (Exercise 24.3); this leaves 6 cases to compute. For example, if $(O' \mid \mathfrak{p}) = *$ and $(O' \mid \mathfrak{p}) = 1$, then

$$\frac{[O'^{\times} : 1+\mathfrak{p}O']}{[O^{\times} : 1+\mathfrak{p}O]} = \frac{q(q-1)^2(q+1)}{q^2(q-1)^2} = 1 + \frac{1}{q}.$$

The other cases follow similarly (Exercise 26.8). □

We can now finish the job.

Proof of Main Theorem 26.1.5. We first invoke Theorem 26.5.4 for a maximal order $O' \supseteq O$ to get

$$\operatorname{mass}(\operatorname{Cls} O') = \frac{2}{(2\pi)^{2n}} h_F d_F^{3/2} \operatorname{Nm}(\mathfrak{D}) \prod_{\mathfrak{p} \mid \mathfrak{D}} \left(1 - \frac{1}{\operatorname{Nm}(\mathfrak{p})}\right).$$

By Proposition 26.6.4 and Lemma 26.6.7, we have

$$\begin{aligned} \operatorname{mass}(\operatorname{Cls} O) &= \operatorname{mass}(\operatorname{Cls} O') \prod_{\mathfrak{p} \mid \mathfrak{N}} [O'^{\times}_{\mathfrak{p}} : O^{\times}_{\mathfrak{p}}] \\ &= \operatorname{mass}(\operatorname{Cls} O')[O' : O]_{\mathbb{Z}} \prod_{\mathfrak{p} \mid \mathfrak{D}} \left(1 - \frac{1}{\operatorname{Nm}(\mathfrak{p})}\right)^{-1} \prod_{\mathfrak{p} \mid \mathfrak{N}} \lambda(O,\mathfrak{p}) \qquad (26.6.8) \\ &= \frac{2}{(2\pi)^{2n}} h_F d_F^{3/2} \operatorname{Nm}(\mathfrak{N}) \prod_{\mathfrak{p} \mid \mathfrak{N}} \lambda(O,\mathfrak{p}) \end{aligned}$$

using $\operatorname{Nm}(\mathfrak{N}) = \operatorname{Nm}(\mathfrak{D})[O' : O]_{\mathbb{Z}}$. □

26.7 Class number one

It is helpful to get a sense of the overall size of the mass, as follows.

26.7.1. Let $m(\mathfrak{D},\mathfrak{M})$ be the mass of an(y) Eichler order of level $\mathfrak{M}$. Then in analogy with the Brauer–Siegel theorem,

$$\log m(\mathfrak{D},\mathfrak{M}) \sim \frac{3}{2}\log d_F + \log h_F + \log \operatorname{Nm}(\mathfrak{D}\mathfrak{M}) \tag{26.7.2}$$

as $d_F \operatorname{Nm}(\mathfrak{D}\mathfrak{M}) \to \infty$ with the degree n fixed: see Exercise 26.9. In particular, for $F = \mathbb{Q}$,

$$\log m(D,M) \sim \log(DM).$$

Since F is totally real, one typically expects h_F to be small in comparison to d_F—but there is a family of real quadratic fields with small regulator first studied by Chowla with $\log h_F \sim \frac{1}{2}\log d_F$, a result due to Montgomery–Weinberger [MW77].

To conclude this section, as in section 25.4 (over $\mathbb{Q}$), the Eichler mass formula can now be used to solve class number one problems for quaternion orders for definite quaternion orders (over totally real fields). This effort was undertaken recently by Kirschmer–Lorch [KL2016]: a complete list of definite orders of type number one is given, and again because $\#\operatorname{Typ} O \le \#\operatorname{Cls} O$, the following theorem can be proven.

Theorem 26.7.3 (Kirschmer–Lorch). *There are* 4194 *one-class genera of primitive, positive definite ternary quadratic forms (equivalently, definite quaternion orders O with $\#\operatorname{Typ} O = 1$, up to isomorphism): they occur over* 30 *possible base fields of degrees up to* 5.

There are exactly 154 *isomorphism classes of definite quaternion orders O with $\#\operatorname{Cls} O = 1$; of these,* 144 *are Gorenstein and* 10 *are non-Gorenstein.*

Remark 26.7.4. Kirschmer–Lorch [KL2016] also enumerate two-class genera; a complete list is available online [KLwww]. Cerri–Chaubert–Lezowski [CCL2013] also consider totally definite Euclidean orders over totally real fields, giving the complete list over $\mathbb{Q}$ and over quadratic fields: all of them are Euclidean under the reduced norm.

26.8 Functional equation and classification

To conclude this chapter, we discuss the functional equation and important applications to the classification of quaternion algebras over number fields. This section serves as preview of the material in chapter 29 (motivation for it!) but we frame results in the same vein as the results of this chapter.

Following Riemann, we complete $\zeta(s) := \sum_{n=1}^{\infty} n^{-s}$ to the function

$$\xi(s) := \pi^{-s/2}\zeta(s)\Gamma(s/2),$$

where $\Gamma(s)$ is the complex Γ-function (Exercise 26.2). Riemann proved that $\xi(s)$ extends to a meromorphic function on $\mathbb{C}$ and satisfies the functional equation

$$\xi(1-s) = \xi(s). \tag{26.8.1}$$

(It is also common to multiply $\xi(s)$ by $s(1-s)$ to cancel the poles at $s = 0, 1$.) We will prove this in section 29.1—the impatient reader is encouraged to flip ahead!

26.8.2. This result extends to Dedekind zeta functions (retaining the notation from section 26.2). Define

$$\Gamma_{\mathbb{R}}(s) := \pi^{-s/2}\Gamma(s/2), \qquad \Gamma_{\mathbb{C}}(s) := 2(2\pi)^{-s}\Gamma(s). \tag{26.8.3}$$

We then define the **completed** Dedekind zeta function to be

$$\xi_F(s) := |d_F|^{s/2}\Gamma_{\mathbb{R}}(s)^r\Gamma_{\mathbb{C}}(s)^c\zeta_F(s). \tag{26.8.4}$$

Then $\xi_F(s)$ satisfies the **functional equation**

$$\xi_F(1-s) = \xi_F(s) \tag{26.8.5}$$

for all $s \in \mathbb{C}$. We prove (26.8.5) as Corollary 29.10.3(a) using idelic methods; this proof will also motivate the completion defined above. For now, we borrow from the future. The functional equation gives $\zeta_F(s)$ meromorphic continuation to $\mathbb{C}$ via

$$\zeta_F(1-s) = \zeta_F(s)\left(\frac{|d_F|}{4^c\pi^n}\right)^{s-1/2}\frac{\Gamma(s/2)^r\Gamma(s)^c}{\Gamma((1-s)/2)^r\Gamma(1-s)^c}. \tag{26.8.6}$$

26.8.7. Using the functional equation (26.8.6), we can rewrite (26.2.4) to obtain the tidier expression

$$\zeta_F^*(0) = \lim_{s\to 0} s^{-(r+c-1)}\zeta_F(0) = \frac{h_F\operatorname{Reg}_F}{w_F}; \tag{26.8.8}$$

in particular, ζ_F has a zero at $s = 0$ of order $r + c - 1$, the rank of the unit group of R by Dirichlet's unit theorem.

In terms of the completed Dedekind zeta function, we find $\xi_F(s)$ has analytic continuation to $\mathbb{C}\smallsetminus\{0, 1\}$ with simple poles at $s = 0, 1$ and residues

$$\xi_F^*(0) = \xi_F^*(1) = \frac{2^{r+c}h_F\operatorname{Reg}_F}{w_F}. \tag{26.8.9}$$

Example 26.8.10. When F is an imaginary quadratic field ($r = 0$ and $c = 1$) we have $\operatorname{Reg}_F = 1$ and $\zeta_F^*(0) = \zeta_F(0)$, so $h_F/w_F = \zeta_F(0)$, and in particular if $|d_F| > 4$ then $h_F = 2\zeta_F(0)$.

We now turn to a quaternionic generalization.

26.8.11. The factorization of $\zeta_B(s)$ in Theorem 26.4.9 implies a functional equation for $\zeta_B(s)$ via the functional equation for $\zeta_F(s)$. This functional equation is simplest to state for a completed zeta function. We recall that

$$\zeta_B(s) = \zeta_F(2s)\zeta_F(2s-1)\prod_{\mathfrak{p}\mid\mathfrak{D}}\left(1 - \operatorname{Nm}(\mathfrak{p})^{1-2s}\right);$$

just as the function ζ_F completes to ξ_F with a simple functional equation, we analogously complete ζ_B to

$$\xi_B(s) := \xi_F(2s)\xi_F(2s-1)\prod_{\mathfrak{p}\mid\mathfrak{D}} \mathrm{Nm}(\mathfrak{p})^s\left(1-\mathrm{Nm}(\mathfrak{p})^{1-2s}\right)\prod_{v\in\Omega}(2s-1) \tag{26.8.12}$$

where $\Omega \subseteq \operatorname{Ram} B$ be the set of real, ramified places in B. The definition (26.8.12) is motivated by the simplicity of the functional equation; see also Remark 26.8.15 below.

Written in a different way,

$$\xi_B(s) = (2\pi)^t(|d_F|^4\,\mathrm{Nm}(\mathfrak{D})^2)^{s/2}\Gamma_B(s)\zeta_B(s) \tag{26.8.13}$$

where t is the number of split real places, so that $\#\,\mathrm{Pl}(F) = \#\Omega + t$,

$$\Gamma_B(s) := \Gamma_{\mathbb{R}}(2s)^r\Gamma_{\mathbb{R}}(2s+1)^{r-t}\Gamma_{\mathbb{R}}(2s-1)^t\Gamma_{\mathbb{C}}(2s)^c\Gamma_{\mathbb{C}}(2s-1)^c, \tag{26.8.14}$$

and we have used the formula

$$\begin{aligned}(2s-1)\Gamma_{\mathbb{R}}(2s-1) &= 2(s-1/2)\pi^{-(2s-1)/2}\Gamma(s-1/2)\\ &= (2\pi)\pi^{-(2s+1)/2}\Gamma(s+1/2)\\ &= 2\pi\Gamma_{\mathbb{R}}(2s+1).\end{aligned}$$

Remark 26.8.15. The completion factors (26.8.12) are not arbitrarily chosen; they have a natural interpretation from an idelic perspective. Perhaps this serves as a motivation for working idelically: namely, that it helps to nail down these kinds of quantities! For more, see section 29.8.

Some properties can be read off easily from (26.8.12).

Proposition 26.8.16 (Analytic continuation, functional equation). *Let $m = \#\operatorname{Ram} B$. Then the following statements hold.*

(a) *$\xi_B(s)$ has meromorphic continuation to $\mathbb{C}$ and is holomorphic in $\mathbb{C}\smallsetminus\{0,1/2,1\}$ with simple poles at $s = 0, 1$.*

(b) *$\xi_B(s)$ satisfies the functional equation*

$$\xi_B(1-s) = (-1)^m\xi_B(s). \tag{26.8.17}$$

(c) *$\xi_B(s)$ has a pole of order $2-m$ at $s = 1/2$; in particular, if $m \geq 2$ then $\xi_B(s)$ is holomorphic at $s = 1/2$.*

Proof. Statement (a) follows from (26.8.12), recalling that $\xi_F(s)$ is holomorphic in $\mathbb{C}\smallsetminus\{0,1\}$ by 26.8.7 with simple poles at $s = 0, 1$. Part (c) follows similarly from (b), since the other factors in (26.8.12) have a simple zero at $s = 1/2$.

To prove (b), we consider each term in the definition of (26.8.12). The functional equation (26.8.5) for $\xi_F(s)$ with $s \leftarrow 1-s$ implies

$$\begin{aligned}\xi_F(2(1-s))\xi_F(2(1-s)-1) &= \xi_F(2-2s)\xi_F(1-2s)\\ &= \xi_F(1-(2-2s))\xi_F(1-(1-2s))\\ &= \xi_F(2s-1)\xi_F(2s).\end{aligned} \tag{26.8.18}$$

For

$$\ell(s) = q^s(1 - q^{1-2s}) = q^s - q^{1-s}$$

and $q > 0$ we have $\ell(1 - s) = -\ell(s)$, so with $q = \mathrm{Nm}(\mathfrak{p})$ the factors $\mathfrak{p} \mid \mathfrak{D}$ are taken into account. Finally, $2(1 - s) - 1 = -(2s - 1)$ takes care of $v \in \Omega$, and (b) follows. □

Proposition 26.8.16 shows how algebraic properties of B correspond to analytic properties of ξ_B. A deeper investigation ultimately will reveal the following fundamental result.

Theorem 26.8.19 (Sign of functional equation, holomorphicity). *$\xi_B(s)$ satisfies the functional equation*

$$\xi_B(1 - s) = \xi_B(s). \tag{26.8.20}$$

Moreover, if B is a division algebra, then $\xi_B(s)$ is holomorphic at $s = 1/2$.

Proof. This theorem was proven by Hey [Hey29, §3] (more generally, for division algebras over $\mathbb{Q}$) following the same general script as in the proof of the functional equation for the Dedekind zeta function (26.8.5), as proven first by Hecke: the key ingredient is Poisson summation. The argument is also given by Eichler [Eic38a, Part V]. We instead prove this theorem in the language of ideles (Main Theorem 29.2.6), as it simplifies the calculations—and so for continuity of ideas in the exposition, we borrow from the future. □

Assuming Theorem 26.8.19, we can now deduce the main classification theorem (Main Theorem 14.6.1) for quaternion algebras over number fields. First, we have Hilbert reciprocity as an immediate consequence.

Corollary 26.8.21 (Hilbert reciprocity, cf. Corollary 14.6.2). *$\#\operatorname{Ram} B$ is even.*

Proof. Immediate from (26.8.17) and (26.8.20). □

Next we conclude the all-important local-global principle.

Corollary 26.8.22. *We have $B \simeq \mathrm{M}_2(F)$ if and only if $B_v \simeq \mathrm{M}_2(F_v)$ for all (but one) places $v \in \operatorname{Pl} F$.*

Proof. The implication (⇒) is immediate. For the converse (⇐), by Proposition 26.8.16(c), $\xi_B(s)$ has a pole of order $2 - m$ at $s = 1/2$, so if $m \leq 1$ then $\xi_B(s)$ is not holomorphic at $s = 1/2$; but then by Theorem 26.8.19, B is not a division algebra, so $B \simeq \mathrm{M}_2(F)$ (and the order of pole is necessarily 2, and $B_v \simeq \mathrm{M}_2(F_v)$ for *all* v). □

From this corollary, we are able to deduce the Hasse norm theorem for quadratic extensions.

Theorem 26.8.23 (Hasse norm theorem). *Let $K \supseteq F$ be a separable quadratic field extension and let $b \in F^\times$. Then $b \in \mathrm{Nm}_{K/F}(K^\times)$ if and only if $b \in \mathrm{Nm}_{K_v/F_v}(K_v^\times)$ for all (but one) places $v \in \operatorname{Pl} F$.*

Proof. Consider the quaternion algebra $B = (K, b \mid F)$. Then by Main Theorem 5.4.4, we have $b \in \mathrm{Nm}_{K|F}(K^\times)$ if and only if $B \simeq \mathrm{M}_2(F)$. By Corollary 26.8.22, this holds if and only if $B_v \simeq \mathrm{M}_2(F_v)$ for all (but one) places v. Repeating the application of Main Theorem 5.4.4, this holds if and only if $B_v \simeq \mathrm{M}_2(F_v)$ for all (but one) v. □

We may similarly conclude the all important local-global principle for quadratic forms.

Theorem 26.8.24 (Hasse–Minkowski theorem). *Let Q be a quadratic form over F. Then Q is isotropic over F if and only if Q_v is isotropic over F_v for all places v of F.*

Proof. The implication ($\Rightarrow$) is immediate, so we prove ($\Leftarrow$). We may suppose without loss of generality that Q is nondegenerate. If $n = \dim_F V = 1$, the theorem is vacuous.

Suppose $n = 2$. Then after scaling we may suppose $Q = \langle 1, -a\rangle$, and Q is isotropic if and only if a is a square. Suppose for purposes of contradiction that $K = F(\sqrt{a})$ is a field. Since Q_v is isotropic for all v, we have $K_v \simeq F_v \times F_v$ for all v, and thus $\zeta_K(s) = \zeta_F(s)^2$. But as Dedekind zeta functions, both $\zeta_F(s)$ and $\zeta_K(s)$ have poles of order 1 at $s = 1$ (we evaluated the residue in the analytic class number formula, Theorem 26.2.3), a contradiction.

Suppose $n = 3$. Again after rescaling we may suppose $Q = \langle 1, -a, -b\rangle$, and Q is isotropic if and only if b is a norm from $F[\sqrt{a}]$: then the equivalence follows from Theorem 26.8.23.

Next, suppose $n = 4$, and $Q = \langle 1, -a, -b, c\rangle$. Let $K = F(\sqrt{abc})$. By extension, Q is isotropic over K and all of its completions. But now $Q \simeq \langle 1, -a, -b, ab\rangle$ over K. Let $B = (a, b \mid K)$. Then by Main Theorem 5.4.4, we have $B_w \simeq \mathrm{M}_2(K_w)$ for all w; thus by Corollary 26.8.22 we have $B \simeq \mathrm{M}_2(K)$, so K splits B. By 5.4.7, we have $K \hookrightarrow B$, so there exist $x, y, z \in F$ such that

$$\mathrm{nrd}(\alpha) = \mathrm{nrd}(xi + yj + zij) = -ax^2 - by^2 + abz^2 = -abc;$$

dividing by ab we have $z^2 - a(y/a)^2 - b(x/b)^2 + c = 0$ so $Q(z, y/a, x/b, 1) = 0$.

Finally, when $n \geq 5$, we make an argument like at the end of proof of Theorem 14.3.3: we follow Lam [Lam2005, Theorem VI.3.8] and Milne [Milne-CFT, Theorem VIII.3.5(b)], but we are brief. Write $Q = Q_1 \perp Q_2$ where $Q_1 = \langle a, b\rangle$ and $\dim_F V_2 \geq 3$. Choosing a ternary subform and looking at its quaternion algebra, we find a finite set $T \subseteq \mathrm{Pl}\, F$ such that Q_2 is isotropic for all $v \notin T$. For each $v \in T$, let $Q(z_v) = 0$ and let $c_v = Q_1(z_v) = -Q_2(z_v)$. Choose $x, y \in F^\times$ close enough so that $z = Q_1(x)$ has $zz_v \in F_v^{\times 2}$. The form $Q' = \langle c\rangle \perp Q_2$ in $n-1$ variables is isotropic for all v: for $v \in T$ this was arranged, and for $v \notin T$ already Q_2 was isotropic at v. By induction on n, we conclude that Q' is isotropic; diagonalizing, we may write $Q = \langle d\rangle \perp Q'$, and it follows that Q is isotropic. □

Corollary 26.8.25. *Let Q, Q' be quadratic forms over F in the same number of variables. Then $Q \simeq Q'$ if and only if $Q_v \simeq Q'_v$ for all places $v \in \mathrm{Pl}\, F$.*

Proof. Apply the same method of proof as in Corollary 14.3.7: see Exercise 26.10. □

We may now conclude the classification with one further input.

Theorem 26.8.26 (Infinitude of primes in arithmetic progression over number fields). *Let* $\mathfrak{n} \subseteq \mathbb{Z}_F$ *be a nonzero ideal, let* $a \in (\mathbb{Z}_F/\mathfrak{n})^\times$*, and for each* $v \mid \infty$ *real let* $\epsilon_v \in \{\pm 1\}$*. Then there are infinitely many prime elements* $p \in \mathbb{Z}_F$ *such that* $\operatorname{sgn}(v(p)) = \epsilon_v$ *and* $p \equiv a \pmod{\mathfrak{n}}$.

Proof. The theorem generalizes Dirichlet's theorem on the infinitude of primes in arithmetic progression (Theorem 14.2.9): see Lang [Lang94, Theorem VIII.4.10] or Neukirch [Neu99, Theorem VII.13.4]. □

Proof of Main Theorem 14.6.1, F a number field. First, the map $B \mapsto \operatorname{Ram} B$ has the correct codomain by Hilbert reciprocity (Corollary 26.8.21). Surjectivity follows by Exercise 14.17 (using Theorem 26.8.26. To conclude, we show injectivity. We refer to Corollary 5.2.6, giving a bijection between quaternion algebras over F up to isomorphism and ternary quadratic forms of discriminant 1 up to isometry; and we recall Theorem 12.3.4, that (rescaling) there is a unique anisotropic ternary quadratic form of discriminant 1 up to isometry. Therefore Corollary 26.8.25 implies that the map $B \mapsto \operatorname{Ram} B$ is injective, since the set $\operatorname{Ram} B$ records those places v where the ternary quadratic form attached to B is anisotropic. □

We will give another proof of Main Theorem 14.6.1 over global fields using the characterization of idelic norms in Proposition 27.5.15 (avoiding fiddling with quadratic forms and the use of primes in arithmetic progression).

Remark 26.8.27. For the readers who accept the fundamental exact sequence of class field theory as in Remark 14.6.10, the arguments above can be run in reverse, and the analytic statement in Theorem 26.8.19 can be deduced as a consequence.

Exercises

1. Prove Proposition 26.2.18 that

$$\operatorname{vol}(X_{\leq 1}) = \frac{2^r (2\pi)^c \operatorname{Reg}_F}{w_F}$$

in the special case of a real quadratic field.

▷ 2. Let

$$\Gamma(s) := \int_0^\infty x^s e^{-x} \frac{\mathrm{d}x}{x}$$

be the complex Γ-function, defined for $\operatorname{Re} s > 0$. Verify the following basic properties of $\Gamma(s)$.

(a) $\Gamma(1) = 1$ and $\Gamma(1/2) = \sqrt{\pi}$.
(b) $\Gamma(s+1) = s\Gamma(s)$ for all $\operatorname{Re} s > 0$, and $\Gamma(n) = (n-1)!$ for $n \geq 1$.
(c) $\Gamma(s)$ has meromorphic continuation to $\mathbb{C}$, holomorphic away from simple poles at $\mathbb{Z}_{\leq 0}$.
(d) $\Gamma(s)$ has no zeros in $\mathbb{C}$.

▷ 3. In this exercise, we prove Proposition 26.2.18. Let F be a number field with ring of integers R, let $X \subseteq F_\mathbb{R}$ be a cone fundamental domain for $R^\times$. Let Reg_F be the regulator of F and w_F the number of roots of unity in F.

(a) Let $V = F_{\mathbb{R}}$ be the ambient space. Let $\mu(R^\times) \leq R^\times$ be the group generated by a fundamental system of units, so $R^\times/\mu(R^\times) \simeq R^\times_{\text{tors}}$. Show that

$$\operatorname{vol}((V/R^\times)_{\leq 1}) = \frac{2^c}{w_F} \int_{V_{\leq 1}/\mu(R^\times)} \mathrm{d}x\, \mathrm{d}z$$

with x_i, z_j standard coordinates on $\mathbb{R}^r \times \mathbb{C}^c$ in multi-index notation.

(b) Let ρ_j, θ_j be polar coordinates on $\mathbb{C}^c$, and restrict the domain V to the domain V^+ with $x_i > 0$ for all i. Let W^+ be the projection of V^+ onto the x, ρ-coordinate plane and let $x_{r+j} = \rho_j^2$. Show that

$$\int_{V_{\leq 1}/\mu(R^\times)} \mathrm{d}x\, \mathrm{d}z = 2^r \pi^c \int_{W^{+,\leq 1}/\mu(R^\times)} \mathrm{d}x.$$

(c) Apply the change of variables $u_i = \log x_i$ to obtain

$$\int_{W^+_{\leq 1}/\mu(R^\times)} \mathrm{d}x = \int_P \mathrm{d}u$$

where P is the fundamental parallelogram for the additive (logarithmic) action of $R^\times$. Conclude that

$$\operatorname{vol}(X_{\leq 1}) = \frac{2^r (2\pi)^c \operatorname{Reg}_F}{w_F}.$$

▹ 4. Show that the ideal J in Lemma 26.3.6 is unique: more specifically, show that if I is an invertible, integral lattice and suppose that $\operatorname{nrd}(I) = \mathfrak{m}\mathfrak{n}$ with $\mathfrak{m}, \mathfrak{n} \subseteq R$ coprime ideals, then an invertible, integral lattice J such that I is compatible with J^{-1} with IJ^{-1} integral and $\operatorname{nrd}(J) = \mathfrak{m}$ is unique.

▹ 5. Let F be a number field with ring of integers R, let B be a quaternion algebra over F, and let $O \subseteq O' \subseteq B$ be R-orders. For $I \subseteq O$ a right O-ideal, we define the right O'-ideal $\rho(I) = IO' \subseteq O'$ obtained by extension.

(a) Show that ρ induces a (well-defined) surjective map

$$\begin{aligned} \operatorname{Cls} O &\to \operatorname{Cls} O' \\ [I] &\mapsto [IO'] \end{aligned}$$

of pointed sets with finite fibers.

(b) For the case where O is the Lipschitz order and O' the Hurwitz order, show that the map in (a) is a bijection (cf. Lemma 11.2.9).

▹ 6. Let R be a DVR with uniformizer π and let I be a (invertible) integral right $\mathrm{M}_2(R)$-ideal. Show that I is generated by

$$x = \begin{pmatrix} \pi^u & 0 \\ c & \pi^v \end{pmatrix}$$

where $u, v \in \mathbb{Z}_{\geq 0}$ and $c \in R/\pi^v$ are unique.

7. Generalize Exercise 11.13 as follows. For $n \in \mathbb{Z}$, let

$$r_4(n) := \#\{(t,x,y,z) \in \mathbb{Z}^4 : t^2 + x^2 + y^2 + z^2 = n\}$$

and let $r_4'(n) := r_4(n)/8$.

   (a) Show that $r_4'(2^e) = 1$ for all $e \geq 1$ and $r_4'(p^e) = 1 + p + \cdots + p^e$ for all $e \geq 1$ and p odd. *[Hint: relate the count to the number of right ideals and inspect the coefficients of the zeta function.]*
   (b) Show that r_4' is a multiplicative function: $r_4'(mn) = r_4'(m)r_4'(n)$ if $\gcd(m,n) = 1$.
   (c) Conclude that

$$r_4(n) = \begin{cases} 8\sum_{d\mid n} d, & n \text{ odd;} \\ 24\sum_{d\mid m} d, & n = 2^e m \text{ even with } m \text{ odd.} \end{cases}$$

▹ 8. Finish the proof of Lemma 26.6.7 by checking the remaining cases.

9. Prove (26.7.2). Specifically, for a number field F and coprime ideals $\mathfrak{D}, \mathfrak{M}$ with $\mathfrak{D}$ squarefree and coprime to $\mathfrak{M}$, define the **mass**

$$m(F,\mathfrak{D},\mathfrak{M}) := \frac{2\zeta_F(2)}{(2\pi)^{2n}} d_F^{3/2} h_F \varphi(\mathfrak{D})\psi(\mathfrak{M}).$$

Let $\mathfrak{N} := \mathfrak{D}\mathfrak{M}$. Show for fixed n that

$$\log m(F,\mathfrak{D},\mathfrak{M}) \sim \frac{3}{2}\log d_F + \log h_F + \log \operatorname{Nm}(\mathfrak{N}) \tag{26.8.28}$$

as $d_F \operatorname{Nm}(\mathfrak{N}) \to \infty$, as follows.

   (a) Show that

$$\zeta_{\mathbb{Q}}(2)^n = \prod_p \left(1 - \frac{1}{p^2}\right)^{-n} \leq \zeta_F(2) \leq \prod_p \left(1 - \frac{1}{p^{2n}}\right)^{-1} = \zeta_{\mathbb{Q}}(2n)$$

   so $\zeta_F(2) \asymp 1$.
   (b) Show that

$$\frac{\operatorname{Nm}(\mathfrak{N})}{\log\log \operatorname{Nm}(\mathfrak{D})} \ll \varphi(\mathfrak{D})\psi(\mathfrak{M}) \ll \operatorname{Nm}(\mathfrak{N})\left(\log\log \operatorname{Nm}(\mathfrak{M})\right).$$

   [Hint: you may need some elementary estimates from analytic number theory, adapted for this purpose; you may wish to start with the case $F = \mathbb{Q}$.]
   (c) Conclude (26.8.28).

10. Prove Corollary 26.8.25: if Q, Q' are quadratic forms over F in the same number of variables, then $Q \simeq Q'$ if and only if $Q_v \simeq Q'_v$ for all places $v \in \operatorname{Pl} F$. *[Hint: see Corollary 14.3.7.]*

▷ 11. Use Dirichlet's analytic class number formula to prove the theorem on arithmetic progressions (Theorem 14.2.9) as follows.

(a) Let $F = \mathbb{Q}(\zeta_m)$. Show that

$$\zeta_F(s) = \zeta(s) \prod_{\chi \neq 1} L(s, \chi)$$

where χ runs over all nontrivial Dirichlet characters $\chi \colon (\mathbb{Z}/m\mathbb{Z})^\times \to \mathbb{C}^\times$, and

$$L(s, \chi) \sum_{n=1}^{\infty} \frac{\chi(n)}{n^s}.$$

[Hint: Factor according to the decomposition $m = ref$.]

(b) Use partial summation and the fact that the partial sums are bounded to show that each $L(s, \chi)$ for $\chi \neq 1$ is holomorphic at $s = 1$.

(c) Conclude from the analytic class number formula that $L(1, \chi) \neq 0$ for $\chi \neq 1$.

(d) For $\gcd(a, m) = 1$, using (c) show that as $s \searrow 1$ that

$$\sum_{p \equiv a \pmod{m}} p^{-s} = \frac{\log \zeta(s)}{\varphi(m)} + O(1)$$

and conclude that the set of primes p with $p \equiv a \pmod{m}$ is infinite.

▷ 12. Let F be a nonarchimedean local field with valuation ring R having maximal ideal $\mathfrak{p}$ and residue field k of size $q := \#k$. For $n \geq 1$, let $B_n := \mathrm{M}_n(F)$ and $O_n := \mathrm{M}_n(R)$. In this exercise we generalize Lemma 26.4.1(b).

(a) Show that the set of right integral O_n-ideals is in bijection with the set

$$\left\{ \begin{pmatrix} \pi^{u_1} & 0 & 0 & \cdots & 0 \\ c_{21} & \pi^{u_2} & 0 & \cdots & 0 \\ c_{31} & c_{32} & \pi^{u_3} & \cdots & 0 \\ \vdots & \vdots & \ddots & \ddots & 0 \\ c_{n1} & c_{n2} & c_{n3} & \cdots & \pi^{u_n} \end{pmatrix} : u_1, \dots, u_n \in \mathbb{Z}_{\geq 0}, c_{ij} \in R/\mathfrak{p}^i \right\}.$$

[Hint: appeal to the theory of elementary divisors, applying column operations acting on the right.]

(b) Let $a_{\mathfrak{p}^e}(O_n)$ be the number of right ideals of reduced norm $\mathfrak{p}^e$ in O. Show that

$$a_{\mathfrak{p}^e}(O_n) = \sum_{f=0}^{e} a_{\mathfrak{p}^f}(O_{n-1}) q^{(n-1)(e-f)}.$$

(c) Let

$$\zeta_{O_n}(s) := \sum_{I \subseteq O_n} \frac{1}{\mathsf{N}(I)},$$

the sum over nonzero right ideals of O_n. Show that $\zeta_{O_1}(s) = \zeta_F(s) = (1 - q^{-s})^{-1}$.

(d) Show for $n \geq 2$ that

$$\zeta_{O_n}(s) = \sum_{e=0}^{\infty} \frac{a_{\mathfrak{p}^e}(O_n)}{q^{nes}} = \zeta_{O_{n-1}}(ns/(n-1))\zeta_{O_1}(ns-(n-1)).$$

(e) Conclude that

$$\zeta_{O_n}(s) = \zeta_F(ns)\zeta_F(ns-1)\cdots\zeta_F(ns-(n-1)) = \prod_{i=0}^{n-1} \zeta_F(ns-i)$$

(cf. Corollary 26.4.7).

This chapter is licensed under the terms of the Creative Commons Attribution-NonCommercial 4.0 International License (http://creativecommons.org/licenses/by-nc/4.0/), which permits any noncommercial use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license and indicate if changes were made.

The images or other third party material in this chapter are included in the chapter's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the chapter's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.

Check for updates

Chapter 27

Adelic framework

We have already seen that the local-global dictionary is a powerful tool in understanding the arithmetic of quaternion algebras. In this section, we formalize this connection by consideration of adeles and ideles.

The basic idea: we want to consider all of the completions of a global field at once. There are at least two benefits to this approach:

- We will gain notational efficiency, resulting in brief and well-behaved proofs that would be difficult or impossible to state clearly in classical language.
- Each completion is a locally compact field and so amenable to harmonic analysis, and by extension to the adele ring and its group of units, we can do harmonic analysis on global objects.

The adelic framework, and its use in class field theory, is a vast topic whose complete development deserves its own book. We do our best in this chapter to develop this notation and state what is needed for the case of quaternion algebras. For further background reading, see Childress [Chi2009] and the references given at the start of section 27.4

27.1 ▹ The rational adele ring

In this first section, we work purely over $\mathbb{Q}$ to give a concrete flavor to the abstract definitions to come.

27.1.1. Recall in section 12.1 that for a prime p we defined $\mathbb{Z}_p = \varprojlim_r \mathbb{Z}/p^r\mathbb{Z}$ as a projective limit, and each $\mathbb{Z}_p$ is compact. We can package these together to make the direct product ring

$$\widehat{\mathbb{Z}} := \prod_p \mathbb{Z}_p \tag{27.1.2}$$

equipped with the product topology: as a profinite group, it is Hausdorff, compact, and totally disconnected.

© The Author(s) 2021

J. Voight, *Quaternion Algebras*, Graduate Texts in Mathematics 288,
https://doi.org/10.1007/978-3-030-56694-4_27

We can see $\widehat{\mathbb{Z}}$ itself as projective limit as follows. By the Sun Zi theorem (CRT) theorem, we have an isomorphism

$$\widehat{\mathbb{Z}} = \prod_p \varprojlim_r \mathbb{Z}/p^r\mathbb{Z} \xrightarrow{\sim} \varprojlim_n \mathbb{Z}/n\mathbb{Z}$$

of topological rings, with the projective limit indexed by positive integers partially ordered under divisibility; so under this isomorphism, we may identify

$$\begin{aligned}\widehat{\mathbb{Z}} &= \varprojlim_n \mathbb{Z}/n\mathbb{Z} \\ &= \left\{(a_n)_{n=1}^{\infty} \in \textstyle\prod_{n=1}^{\infty} \mathbb{Z}/n\mathbb{Z} : a_m \equiv a_n \pmod{n} \text{ for all } n \mid m\right\}\end{aligned} \tag{27.1.3}$$

The natural ring homomorphism $\mathbb{Z} \to \widehat{\mathbb{Z}}$ which takes every element to its reduction modulo n is injective; the image of $\mathbb{Z}$ is discrete and dense in $\widehat{\mathbb{Z}}$ again by the CRT. One warning is due: $\widehat{\mathbb{Z}}$ is not a domain.

27.1.4. We now make the ring $\widehat{\mathbb{Z}}$ a bit bigger so that it contains $\mathbb{Q}$ as a subring. If we were to take the ring $\prod_p \mathbb{Q}_p$, a product of locally compact rings, unfortunately we would no longer have something that is locally compact (see Exercise 27.1): the product $\prod_p \mathbb{Q}_p$ is much too big, allowing denominators in every component, whereas the image of $\mathbb{Q}$ will only have denominators in finitely many positions. We should also keep track of archimedean information at the same time.

With these in mind we define, for each finite set S of primes, the ring

$$U_S := \mathbb{R} \times \prod_{p \in S} \mathbb{Q}_p \times \prod_{p \notin S} \mathbb{Z}_p \tag{27.1.5}$$

equipped with the product topology, so that U_S is locally compact. For example,

$$U_\emptyset = \mathbb{R} \times \widehat{\mathbb{Z}}. \tag{27.1.6}$$

To assemble these rings together, allowing more denominators and arbitrarily large sets S, we take the injective limit of U_S under the natural directed system $U_S \hookrightarrow U_{S'}$ for $S \subseteq S'$. The resulting object is the **restricted direct product** of $\mathbb{Q}_p$ relative to $\mathbb{Z}_p$ and is called the **adele ring** $\underline{\mathbb{Q}}$ of $\mathbb{Q}$:

$$\begin{aligned}\underline{\mathbb{Q}} &:= \mathbb{R} \times \prod_p{}' \mathbb{Q}_p = \prod_{v \in \operatorname{Pl}\mathbb{Q}}{}' \mathbb{Q}_v \\ &:= \mathbb{R} \times \left\{\underline{x} = (x_p)_p \in \textstyle\prod_p \mathbb{Q}_p : x_p \in \mathbb{Z}_p \text{ for all but finitely many } p\right\} \\ &= \left\{\underline{x} = (x_v)_v \in \textstyle\prod_v \mathbb{Q}_v : |x_v|_v \leq 1 \text{ for all but finitely many } v\right\}\end{aligned} \tag{27.1.7}$$

We declare the sets $U_S \subseteq \underline{\mathbb{Q}}$ with the product topology to be open in $\underline{\mathbb{Q}}$; and with this basis of open neighborhoods of 0 (open in U_S for some S), we have given $\underline{\mathbb{Q}}$ the structure of a topological ring. The sets $U_S \subseteq \underline{\mathbb{Q}}$ are also closed. Note that the topology on $\underline{\mathbb{Q}} \subset \prod_v \mathbb{Q}_v$ is *not* the subspace topology.

27.1.8. For a finite set S, we write

$$\underline{\mathbb{Q}}_{\not S} := \prod_{v \notin S}{}' \mathbb{Q}_v \tag{27.1.9}$$

for the projection of $\underline{\mathbb{Q}}$ onto the factors away from S. We also write

$$\mathbb{Q}_S := \prod_{v \in S} \mathbb{Q}_v. \tag{27.1.10}$$

We embed each of these into $\underline{\mathbb{Q}}$ extending by zero and identify them with their images, so that $\underline{\mathbb{Q}} = \underline{\mathbb{Q}}_{\not S} \times \mathbb{Q}_S$.

Remark 27.1.11. Our notation $\underline{\mathbb{Q}}$ for the adele ring returns to the notation of Weil [Weil82] but is not standard; more typically, the adele ring is denoted $\mathbb{A}$ (which we find markedly problematic).

Remark 27.1.12. Although $\underline{\mathbb{Q}}_{\not S}$ and $\mathbb{Q}_S$ are rings and (via projection) are naturally quotient rings of $\underline{\mathbb{Q}}$, they are *not* subrings because they do not contain 1. This subtlety should cause no confusion in what follows (especially because we will be focused on the multiplicative case and working with groups, where there is no issue extending by the multiplicative identity 1).

We have a natural continuous embedding $\mathbb{Q} \hookrightarrow \mathbb{Q}_v$ for all $v \in \operatorname{Pl}\mathbb{Q}$, and this extends to a diagonal embedding $\mathbb{Q} \hookrightarrow \underline{\mathbb{Q}}$.

Lemma 27.1.13. *The diagonal embedding $\mathbb{Q} \hookrightarrow \underline{\mathbb{Q}}$ is a continuous injective ring homomorphism and the image is closed and discrete as a subring of $\underline{\mathbb{Q}}$.*

Proof. The embedding is continuous because it is so in each component. Because $\underline{\mathbb{Q}}$ is a topological group under addition, to prove the remaining part it is enough to find a neighborhood $0 \in U \subseteq \underline{\mathbb{Q}}$ such that $U \cap \mathbb{Q} = \{0\}$. We take

$$U := (-1,1) \times \widehat{\mathbb{Z}} = \left\{(x_v)_v : |x_\infty|_\infty < 1 \text{ and } |x_p|_p \le 1 \text{ for all primes } p\right\}.$$

By definition, U is open in $\underline{\mathbb{Q}}$ as it is open in $U_\emptyset$ (for a reminder, see (27.1.6)). And if $a \in U \cap \mathbb{Q}$, then $a \in \mathbb{Z}_p$ for all p, so $a \in \mathbb{Z}$, and $|a|_\infty < 1$, and thus $a = 0$. ◻

Lemma 27.1.14. *The image of $\mathbb{Q} \hookrightarrow \underline{\mathbb{Q}}$ is cocompact, i.e., $\underline{\mathbb{Q}}/\mathbb{Q}$ is compact.*

Proof. Let $W := [0,1] \times \widehat{\mathbb{Z}}$. Then W is compact. By strong approximation—for a snapshot review, flip ahead to Theorem 28.1.9 and its corollary—we have $\underline{\mathbb{Q}} = \mathbb{Q} + W$. Therefore the continuous quotient map $\underline{\mathbb{Q}} \to \underline{\mathbb{Q}}/\mathbb{Q}$ restricted to W is surjective, so $\underline{\mathbb{Q}}/\mathbb{Q}$ equal to the image of the compact set W is compact. (W is a fundamental set for the action of $\mathbb{Q}$ on $\underline{\mathbb{Q}}$: see Exercise 27.2.) ◻

27.1.15. The proof of Lemma 27.1.14 shows that the natural map $\mathbb{R} \times \widehat{\mathbb{Z}} \to \underline{\mathbb{Q}}/\mathbb{Q}$ is surjective; its kernel is $\mathbb{Z}$ diagonally embedded, so we have an isomorphism

$$\underline{\mathbb{Q}}/\mathbb{Q} \xrightarrow{\sim} (\mathbb{R} \times \widehat{\mathbb{Z}})/\mathbb{Z}$$

of topological groups. The resulting topological group $\mathrm{Sol} := (\mathbb{R} \times \widehat{\mathbb{Z}})/\mathbb{Z}$ is called a **solenoid**: it is compact, Hausdorff, connected, but not path-connected (Exercise 27.5), which can be visualized as in Figure 27.1.16.

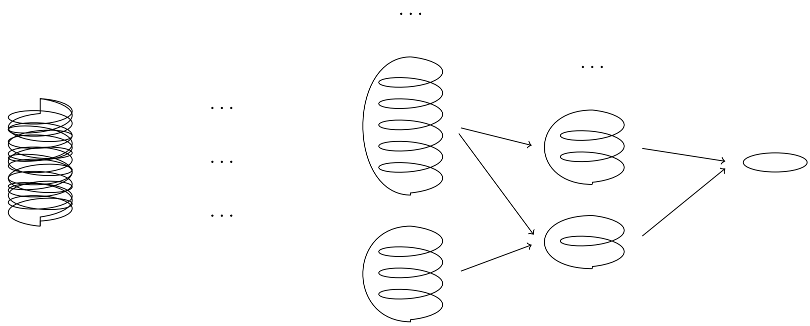


Figure 27.1.16: The solenoid

Very often, we will want to tease apart the nonarchimedean and archimedean parts of the adele ring $\underline{\mathbb{Q}}$, and will write

$$\widehat{\mathbb{Q}} := \underline{\mathbb{Q}}_{\{\infty\}} = \prod_p{}' \mathbb{Q}_p \simeq \widehat{\mathbb{Z}} \otimes_{\mathbb{Z}} \mathbb{Q} \tag{27.1.17}$$

so that extending by zero we have $\underline{\mathbb{Q}} = \widehat{\mathbb{Q}} \times \mathbb{R}$.

27.2 ▹ The rational idele group

Having dealt with the additive version in adeles, now we talk about the multiplicative version, ideles.

27.2.1. We define the **rational idele group** to be

$$\begin{aligned} \underline{\mathbb{Q}}^\times &= \mathbb{R}^\times \times \prod_p{}' \mathbb{Q}_p^\times = \mathbb{R}^\times \times \widehat{\mathbb{Q}}^\times \\ &= \left\{ \underline{x} = (x_v)_v \in \textstyle\prod_v \mathbb{Q}_v^\times : |x_v| = 1 \text{ for all but finitely many } v \right\} \end{aligned} \tag{27.2.2}$$

That is to say, $\underline{\mathbb{Q}}^\times$ is the restricted direct product of the spaces $\mathbb{Q}_p^\times$ with respect to $\mathbb{Z}_p^\times$. The topology is such that for S a finite set of primes, the set

$$V_S = \mathbb{R}^\times \times \prod_{p \in S} \mathbb{Q}_p^\times \times \prod_{p \notin S} \mathbb{Z}_p^\times$$

is open and closed as a subgroup of $\underline{\mathbb{Q}}^\times$.

Remark 27.2.3. Chevalley first used the name *élément idéal* for elements of $\underline{\mathbb{Q}}^\times$, but at Hasse's suggestion he abbreviated it to *idèle*; the name *adèle* was then shorthand for an "*additive* idele". Anglifying, we drop the accents on these words.

27.2.4. The topology on $\underline{\mathbb{Q}}^\times$ is *not* the subspace topology $\underline{\mathbb{Q}}^\times \subset \underline{\mathbb{Q}}$, because inversion need not be continuous. Instead, we think of $\underline{\mathbb{Q}}^\times$ as a subset of $\underline{\mathbb{Q}} \times \underline{\mathbb{Q}}$ via the map $x \mapsto (x, x^{-1})$, and then $\underline{\mathbb{Q}}^\times$ inherits the structure of a topological group.

Lemma 27.2.5. *The diagonal map $\mathbb{Q}^\times \hookrightarrow \underline{\mathbb{Q}}^\times$ is an injective continuous group homomorphism, and the image of $\mathbb{Q}^\times$ is closed and discrete.*

Proof. Since $\mathbb{Q}$ is closed and discrete in $\underline{\mathbb{Q}}$ and $\underline{\mathbb{Q}}^\times \subseteq \underline{\mathbb{Q}} \times \underline{\mathbb{Q}}$ has the subspace topology, so too is $\mathbb{Q}^\times$ closed and discrete. □

27.2.6. We now give an explicit description of the quotient $\underline{\mathbb{Q}}^\times/\mathbb{Q}^\times$: we will see it is *not* compact.

There is a canonical isomorphism of topological groups

$$\mathbb{Q}_p^\times \simeq \langle p \rangle \times \mathbb{Z}_p^\times$$

by p-adic valuation. Since $\langle p \rangle = p^{\mathbb{Z}} \simeq \mathbb{Z}$, we have a topological group isomorphism

$$\underline{\mathbb{Q}}^\times = \mathbb{R}^\times \times \prod_p{}' \mathbb{Q}_p^\times \simeq \{\pm 1\} \times \mathbb{R}_{>0} \times \prod_p \mathbb{Z}_p^\times \times \bigoplus_p \mathbb{Z}. \tag{27.2.7}$$

A direct sum appears because an element of the restricted direct product is a p-adic unit for all but finitely many p. We project $\underline{\mathbb{Q}}^\times$ onto the product of the first and last factor, getting a continuous surjective map

$$\underline{\mathbb{Q}}^\times \to \{\pm 1\} \times \bigoplus_p \mathbb{Z}. \tag{27.2.8}$$

Looking at $r \in \mathbb{Q}^\times \subseteq \underline{\mathbb{Q}}^\times$, if we write $r = \epsilon \prod_p p^{n(p)}$, where $\epsilon \in \{\pm 1\}$ and $n(p) = \operatorname{ord}_p(r)$, then $r \mapsto (\epsilon, (n(p))_p)$ in the projection. Therefore $\mathbb{Q}^\times$ is canonically identified with $\{\pm 1\} \times \bigoplus_p \mathbb{Z}$ in $\underline{\mathbb{Q}}^\times$. So the projection map (27.2.8) restricts to an isomorphism on the diagonally embedded $\mathbb{Q}^\times$. Therefore

$$\underline{\mathbb{Q}}^\times \simeq \mathbb{Q}^\times \times \mathbb{R}_{>0} \times \prod_p \mathbb{Z}_p^\times. \tag{27.2.9}$$

By the logarithm map, there is an isomorphism $\mathbb{R}_{>0} \simeq \mathbb{R}$, so

$$\underline{\mathbb{Q}}^\times \simeq \mathbb{Q}^\times \times \mathbb{R} \times \widehat{\mathbb{Z}}^\times \tag{27.2.10}$$

and we have an isomorphism of topological groups

$$\underline{\mathbb{Q}}^\times/\mathbb{Q}^\times \simeq \mathbb{R} \times \widehat{\mathbb{Z}}^\times. \tag{27.2.11}$$

(This is *not* a solenoid!)

In a similar way, we see that $\widehat{\mathbb{Q}}^\times/\mathbb{Q}_{>0}^\times \simeq \widehat{\mathbb{Z}}^\times$, where $\mathbb{Q}_{>0}^\times = \{x \in \mathbb{Q} : x > 0\}$, and so $\widehat{\mathbb{Q}}^\times/\mathbb{Q}^\times$ is compact.

Remark 27.2.12. In 27.2.6 we used that $\mathbb{Z}$ is a UFD and $\mathbb{Z}^\times = \{\pm 1\}$; for a general number field, we face problems associated with units and the class group of the field, and the relevant exact sequences will not split!

27.3 ▹ Rational quaternionic adeles and ideles

In the remainder of this chapter, we generalize the above construction to the adele ring and idele group of a global field and then a quaternion algebra over a global field. For the reader on a brisk read, in this section we briefly consider the constructions for a quaternion algebra over $\mathbb{Q}$.

Let B be a quaternion algebra over $\mathbb{Q}$, and let $O \subset B$ be an order.

27.3.1. The **adele ring** of B is

$$\underline{B} := \prod_v{}' B_v = \left\{\alpha = (\alpha_v)_v \in \textstyle\prod_{v \in \mathrm{Pl}\,\mathbb{Q}} B_v : \alpha_v \in O_v \text{ for all but finitely many } v\right\}, \tag{27.3.2}$$

the restricted direct product of the topological rings B_v with respect to O_v for places v of $\mathbb{Q}$; this definition is independent of the choice of order O, because any two orders are equal at all but finitely many places by the local-global dictionary for lattices (Theorem 9.4.9). We embed $B \hookrightarrow \underline{B}$ diagonally: the image is discrete, closed, and cocompact since the same is true for $\mathbb{Q} \hookrightarrow \underline{\mathbb{Q}}$ (Lemmas 27.1.13–27.1.14). We write

$$\widehat{B} := \prod_p{}' B_p, \tag{27.3.3}$$

so extending by zero we may identify $\underline{B} = \widehat{B} \times B_\infty$. We also define

$$\widehat{O} := \prod_p O_p \subseteq \widehat{B}. \tag{27.3.4}$$

The **idele group** of B is $\underline{B}^\times := \prod_v' B_v^\times$, the restricted direct product of the topological groups $B_v^\times$ with respect to $O_v^\times$; we similarly define

$$\widehat{O}^\times := \prod_p O_p^\times \leq \widehat{B}^\times := \prod_p{}' B_p^\times. \tag{27.3.5}$$

Working adelically is notationally quite convenient, as the following lemma illustrates (Lemma 27.6.8 for $F = \mathbb{Q}$).

Lemma 27.3.6. *The set of invertible right fractional O-ideals is in bijection with $\widehat{B}^\times/\widehat{O}^\times$ via the map $I \mapsto \widehat{\alpha}\widehat{O}^\times$, where $I_p = \alpha_p O_p$ and $\widehat{\alpha} = (\alpha_p)_p$; this map induces a bijection*

$$\begin{aligned} \operatorname{Cls}_{\mathsf{R}} O &\leftrightarrow B^\times\backslash\widehat{B}^\times/\widehat{O}^\times \\ [I]_{\mathsf{R}} &\mapsto B^\times\widehat{\alpha}\widehat{O}^\times. \end{aligned}$$

The most fundamental result in this chapter is the following (see Main Theorem 27.6.14, taking $F = \mathbb{Q}$).

Theorem 27.3.7. *Let B be a division quaternion algebra over $\mathbb{Q}$. Then $B^\times \leq \underline{B}^\times$ is cocompact and the set $B^\times\backslash\widehat{B}^\times/\widehat{O}^\times$ is finite.*

In particular, combining Lemma 27.3.6 and Theorem 27.3.7, we conclude that the class set of O is finite, something we proved earlier using the geometry of numbers for B definite in section 17.5.

27.4 Adeles and ideles

In pursuit of more generality, we now repeat the constructions of adeles and ideles over a global field. For further reference, see e.g. Cassels [Cas2010], Goldstein [Gol71], Knapp [Kna2016, Chapter VI], Lang [Lang94, Chapter VII], Neukirch [Neu99, Chapter VI], or Ramakrishnan–Valenza [RM99, Chapter 5].

Throughout the rest of this chapter, let F be a global field.

27.4.1. We recall notation from section 14.4 for convenience. The set of places of F is denoted $\operatorname{Pl} F$. For a place v of F, we denote by F_v the completion of F at the place v, with preferred (normalized) absolute value $|\,|_v$ so that the product formula holds in F: see 14.4.12. If v is nonarchimedean, we let

$$R_v := \{x \in F_v : |x|_v \leq 1\} = \{x \in F_v : v(x) \geq 0\} \tag{27.4.2}$$

be the valuation ring of F_v, where we write v also for the discrete valuation associated to the place v. If F is a number field, we will sometimes denote an archimedean place by writing $v \mid \infty$, and for an archimedean place we just take $R_v = F_v$. A set $S \subseteq \operatorname{Pl} F$ of places is *eligible* if it is finite, nonempty, and contains all archimedean places.

27.4.3. The **adele ring** of F is the restricted direct product of F_v with respect to R_v:

$$\begin{aligned} \underline{F} &:= \prod_v{}' F_v \\ &:= \left\{(x_v)_v \in \textstyle\prod_v F_v : x_v \in R_v \text{ for all but finitely many } v\right\} \\ &= \left\{(x_v)_v \in \textstyle\prod_v F_v : |x_v|_v \leq 1 \text{ for all but finitely many } v\right\} \end{aligned} \tag{27.4.4}$$

with the restricted direct product topology. The topology is uniquely characterized (as a topological ring) by the condition that $\underline{R} := \prod_v R_v$ (with the product topology) is open. Accordingly, a subset $U \subseteq \underline{F}$ is open if and only if for all $\underline{a} \in \underline{F}$, the set $(\underline{a} + U) \cap \prod_v R_v$ is open in the product topology.

Giving F the discrete topology, we have an isomorphism of topological rings $\underline{\mathbb{Q}} \otimes_{\mathbb{Q}} F \xrightarrow{\sim} \underline{F}$: see Exercise 27.14.

27.4.5. We embed $F \subseteq \underline{F}$ under the product of the embeddings $F \hookrightarrow F_v$, i.e., by $x \mapsto (x)_v$; this map is well-defined because $|x|_v \leq 1$ for all but finitely many places v of F. The image of F in $\underline{F}$ has the discrete topology and is closed in $\underline{F}$; the quotient $\underline{F}/F$ is compact (i.e., F is cocompact in $\underline{F}$): for $\underline{x} \in \underline{F}$, find $a \in F$ such that $x_v - a_v \in R_v$ for all v by the Sun Zi theorem (CRT).

27.4.6. Let $S \subset \operatorname{Pl} F$ be an eligible set. We will write

$$\underline{F}_{\not S} := \prod_{\mathfrak{p} \notin S}{}' F_{\mathfrak{p}}, \qquad F_S := \prod_{v \in S} F_v \tag{27.4.7}$$

and extending by zero we identify these sets with their images in $\underline{F}$, so that $\underline{F} = \underline{F}_{\not S} \times F_S$; we call $\underline{F}_{\not S}$ the **S-finite adele ring** of F.

We pass now to the multiplicative situation.

27.4.8. The **idele group** of F is the restricted direct product of $F_v^\times$ with respect to $R_v^\times$:

$$\begin{aligned} \underline{F}^\times &:= \prod_v{}' F_v^\times \\ &:= \left\{(x_v)_v \in \textstyle\prod_v F_v^\times : |x_v|_v = 1 \text{ for all but finitely many } v\right\}. \end{aligned} \tag{27.4.9}$$

27.4.10. The topology on $\underline{F}^\times$ (as a topological ring) is uniquely characterized by the condition that $\prod_v R_v^\times$ (with the product topology) is open. Thus, $U \subseteq \underline{F}^\times$ is open if and only if for all $a \in \underline{F}^\times$, the set $aU \cap \prod_v R_v^\times$ is open in the product topology.

Note that $\underline{F}^\times$ does not have the topology induced from being a subspace of $\underline{F}$, since inversion is not a continuous operation. In general, if A is a topological ring, $A^\times$ becomes a topological group when $A^\times$ is given the relative topology from

$$\begin{aligned} A^\times &\hookrightarrow A \times A \\ x &\mapsto (x, x^{-1}). \end{aligned}$$

(See Exercise 27.13.)

Just as $F \subseteq \underline{F}$ is discrete, $F^\times \subseteq \underline{F}^\times$ is also discrete.

Definition 27.4.11. The group $C_F := \underline{F}^\times / F^\times$ is the **idele class group** of F.

The justification for calling this the idele class group is given in section 27.5.

27.4.12. As above, if $S \subset \operatorname{Pl} F$ is an eligible set, we define the ***S*-finite ideles**

$$\underline{F}_{\not S}^\times := \prod_{v \notin S}{}' F_v^\times$$

which is missing the product at the places in S (a finite product), namely

$$F_S^\times := \prod_{v \in S} F_v^\times;$$

extending by 1, we identify these with their images in $\underline{F}^\times$, so that $\underline{F}^\times = (\underline{F}_{\not S})^\times \times F_S^\times$.

27.4.13. With respect to the normalized absolute values 14.4.12, we have a natural map

$$\begin{aligned} \underline{F}^\times &\to \mathbb{R}_{>0} \\ (x_v)_v &\mapsto \prod_v \|x_v\|_v. \end{aligned} \tag{27.4.14}$$

When F is a number field, the map (27.4.14) is surjective; when F is a function field with constant field $\mathbb{F}_q$, the image is $q^{\mathbb{Z}}$, the cyclic subgroup of $\mathbb{R}_{>0}$ generated by q. Let

$$\underline{F}^{(1)} := \{\underline{x} = (x_v)_v : \textstyle\prod_v \|x_v\|_v = 1\} \tag{27.4.15}$$

so that $\underline{F}^{(1)}$ is the kernel of (27.4.14). Then $F^\times \le \underline{F}^{(1)}$ by the product formula (14.4.6).

The following theorem is fundamental.

Theorem 27.4.16. *The quotient $\underline{F}^{(1)}/F^\times$ is compact, i.e., $F^\times$ is cocompact in $\underline{F}^{(1)}$.*

Proof. We give a proof in a more general context in Main Theorem 27.6.14 below. Or see e.g. Cassels [Cas2010, §16, p. 69] for a direct proof. □

Theorem 27.4.16 is equivalent (!) to the Dirichlet unit theorem and the finiteness of the class group in the number field case, and finite generation of the unit group of a coordinate ring of a curve and the finiteness of the group of rational divisors of degree zero in the function field case [Cas2010, §§17–18].

Via the projection map $\underline{F}^{(1)} \to \underline{F}_{\not S}^\times$, we have $F^\times$ cocompact also in $\underline{F}_{\not S}^\times$.

27.5 Class field theory

Let F^{sep} be a separable closure of F. In this chapter, we summarize the idelic approach to class field theory; unfortunately, we must omit most proofs, as a full treatment would require a lengthy development—but the reader who is willing to accept the statements should be able to digest what follows and will hopefully be motivated to dig deeper! For further reading, see Tate [Tate2010], Lang [Lang94, Chapters XI-XI], Neukirch [Neu99, Chapters IV–VI], or Janusz [Jan96, Chapter V].

27.5.1. Let $R = R_{(S)}$ be a global ring (the ring of S-integers) for the eligible set $S \subseteq \operatorname{Pl} F$. Then R is a Dedekind domain with field of fractions F. The class group of R admits an idelic description, embodying the definitions above, as follows.

To simplify notation, throughout we abbreviate $\underline{F}_{\not S} = \widehat{F}$, as we take the set S to be fixed. To an invertible fractional ideal $\mathfrak{a} \subseteq F$ of R, we have $\mathfrak{a}_\mathfrak{p} = R_\mathfrak{p}$ for all but finitely many primes $\mathfrak{p}$, so we can consider its idelic image $(\mathfrak{a}_\mathfrak{p})_\mathfrak{p} \subseteq \widehat{F}$ under the product of completions. Since $\mathfrak{a}$ is locally principal, we can write each $\mathfrak{a}_\mathfrak{p} = a_\mathfrak{p} R_\mathfrak{p}$ with $a_\mathfrak{p} \in F_\mathfrak{p}^\times$, well-defined up to an element of $R_\mathfrak{p}^\times$; putting these together we obtain an element

$$\widehat{a} = (a_\mathfrak{p})_\mathfrak{p} \in \widehat{F}^\times$$

and

$$\widehat{\mathfrak{a}} = \mathfrak{a}\widehat{R} = \widehat{a}\widehat{R} \subseteq \widehat{F}.$$

We recover $\mathfrak{a} = \widehat{\mathfrak{a}} \cap F$ from Lemmas 9.4.6 and 9.5.3. Therefore the group of invertible fractional ideals of R is canonically isomorphic to the quotient

$$\operatorname{Idl} R \simeq \widehat{F}^\times/\widehat{R}^\times. \tag{27.5.2}$$

The principal (invertible) fractional ideals correspond to the image of $F^\times$ in $\widehat{F}^\times$. Therefore there is a canonical isomorphism

$$\begin{aligned} \operatorname{Cl} R &\xrightarrow{\sim} \widehat{F}^\times/\widehat{R}^\times F^\times \\ [\mathfrak{a}] &\mapsto \widehat{a}\widehat{R}^\times F^\times. \end{aligned} \tag{27.5.3}$$

27.5.4. Suppose F is a number field. If S consists of the set of archimedean places, then $R = \mathbb{Z}_F$ is the ring of integers, and $\operatorname{Cl} R$ is the usual class group. For larger sets S, we have a natural quotient map $\operatorname{Cl} \mathbb{Z}_F \to \operatorname{Cl} R$ obtained by the quotient by the classes of primes $\mathfrak{p}$ corresponding to nonarchimedean places in S.

More generally, one may restrict (27.5.3) to the subgroup of principal fractional ideals which have a totally positive generator; we then obtain the **narrow** (or **strict**) ***S*-class group**

$$\operatorname{Cl}^+ R \xrightarrow{\sim} \widehat{F}^\times/\widehat{R}^\times F_{>0}^\times. \tag{27.5.5}$$

In idelic class field theory, it is often convenient to move between quotients of the finite ideles and quotients of the full idele group as follows.

Lemma 27.5.6. *Let* $F_\infty := \prod_{v\mid\infty} F_v$ *and let*

$$F_{\infty,>0}^\times := \{(a_v)_v \in F_\infty : a_v > 0 \text{ for all v real}\}.$$

Then the map

$$\begin{aligned} \widehat{F}^\times &\to \underline{F}^\times \simeq \widehat{F}^\times \times F_\infty^\times \\ \widehat{a} &\mapsto (\widehat{a}, 1) \end{aligned} \tag{27.5.7}$$

induces isomorphisms of topological groups

$$\begin{aligned} \widehat{F}^\times/F^\times &\xrightarrow{\sim} \underline{F}^\times/F^\times F_\infty^\times \\ \widehat{F}^\times/F_{>0}^\times &\xrightarrow{\sim} \underline{F}^\times/F^\times F_{\infty,>0}^\times. \end{aligned} \tag{27.5.8}$$

Proof. Composing with the projection to $\underline{F}^\times/F^\times F_{\infty,>0}^\times$, we see that the kernel is $F_{>0}^\times$ and that the induced map is surjective; similarly for the quotient by $F^\times F_\infty^\times$. The few details are requested in Exercise 27.9. ◻

Using Lemma 27.5.6, we have

$$\begin{aligned} \operatorname{Cl} R &\xrightarrow{\sim} \underline{F}^\times/(\widehat{R}^\times F_\infty^\times F^\times) \\ \operatorname{Cl}^+ R &\xrightarrow{\sim} \underline{F}^\times/(\widehat{R}^\times F_{\infty,>0}^\times F^\times). \end{aligned}$$

27.5.9. Class field theory relates class groups to abelian extensions. For example, let

$$H = \widehat{R}^\times \times F_S^\times = \prod_{v\notin S} R_v^\times \times \prod_{v\in S} F_v^\times \leq C_F := \underline{F}^\times/F^\times.$$

Then $H \leq C_F$ is an open subgroup of finite index, and the projection map

$$C_F/H \xrightarrow{\sim} \widehat{F}^\times/\widehat{R}^\times$$

is an isomorphism, which together with (27.5.3) gives an isomorphism to $\operatorname{Cl} R$. So we are led to consider the finite-index open subgroups of C_F.

The main theorem of idelic class field theory for finite extensions is as follows.

Theorem 27.5.10. *There is a bijection*

$$\begin{aligned} \{K \subseteq F^{\text{sep}} : K \supseteq F \textit{ finite abelian}\} &\leftrightarrow \{H \le C_F : H \textit{ finite-index open}\} \\ K &\mapsto F^\times \operatorname{Nm}_{K/F} C_K \end{aligned} \tag{27.5.11}$$

together with functorial isomorphisms $C_F/H \xrightarrow{\sim} \operatorname{Gal}(K \mid F)$.

The map $C_F/H \xrightarrow{\sim} \operatorname{Gal}(K \mid F)$ is called the **Artin isomorphism** for H, K.

Proof. See e.g. Tate [Tate2010, §5]. ◻

27.5.12. Rewriting the main theorem (Theorem 27.5.10) slightly, we see that if $H \le \widehat{F}^\times$ is an open finite-index subgroup containing $F^\times_{>0}$, then there is a finite abelian extension $K \supseteq F$ with the Artin isomorphism

$$\widehat{F}^\times/H \xrightarrow{\sim} \operatorname{Gal}(K \mid F).$$

27.5.13. Combining the surjections $C_F \to \operatorname{Gal}(K \mid F)$, we obtain a continuous homomorphism

$$\theta \colon C_F \to \varprojlim_K \operatorname{Gal}(K \mid F) = \operatorname{Gal}(F^{\text{ab}} \mid F)$$

called the **global Artin homomorphism**, where $F^{\text{ab}} \subseteq F^{\text{sep}}$ is the maximal abelian extension of F in F^{sep}.

If F is a number field, then θ is surjective; let D_F be the connected component of 1 in C_F. Then D_F is a closed subgroup with

$$D_F \simeq \mathbb{R} \times (\mathbb{R}/\mathbb{Z})^c \times \operatorname{Sol}^{r+c-1} \tag{27.5.14}$$

(see Exercise 27.10). We therefore have an isomorphism $C_F/D_F \simeq \operatorname{Gal}(F^{\text{ab}} \mid F)$.

If F is a function field with finite constant field k, then θ is injective, and $\theta(C_F)$ is the dense subgroup of automorphisms $\sigma \in \operatorname{Gal}(F^{\text{ab}} \mid F)$ whose restriction to $\operatorname{Gal}(k^{\text{al}} \mid k) \simeq \widehat{\mathbb{Z}}$ lies in $\mathbb{Z}$, i.e., acts by an integer power of the Frobenius. See Tate [Tate2010, §5.4–5.7].

We conclude with a nice application to the classification of quaternion algebras.

Proposition 27.5.15. *Let* $\Sigma \subseteq \operatorname{Pl} F$ *be a finite subset of noncomplex places of* F *of even cardinality. Then there exists a quaternion algebra* B *over* F *with* $\operatorname{Ram} B = \Sigma$.

Proof. Let $K \supseteq F$ be a separable quadratic extension that is inert (an unramified field extension) at every $v \in \Sigma$: such an extension exists by Exercise 14.21. By the main theorem of class field theory, we have $[C_F : F^\times \operatorname{Nm}_{K/F} C_K] = [K : F] = 2$, where $C_F = \underline{F}^\times/F^\times$ and similarly C_K are idele class groups. Therefore

$$[C_F : F^\times \operatorname{Nm}_{K/F} C_K] = [\underline{F}^\times : F^\times \operatorname{Nm}_{K/F}(\underline{K}^\times)] = 2 \tag{27.5.16}$$

as well.

For each $v \in \Sigma$, let π_v be a uniformizer for R_v and if v is real let $\pi_v = -1$. Since $K_v \supseteq F_v$ is an unramified field extension, we have $\pi_v \notin \operatorname{Nm}_{K_v/F_v}(K_v^\times)$. For $v \in \Sigma$, let $\underline{\pi_v} = (1, \dots, 1, \pi_v, \dots) \in \underline{F}^\times$. Then $\underline{\pi_v} \notin \operatorname{Nm}_{K/F}(\underline{K}^\times)$.

We claim that $\underline{\pi_v} \notin F^\times \operatorname{Nm}_{K/F}(\underline{K}^\times)$. Otherwise, there would exist $a \in F^\times$ such that $a\underline{\pi_v} \in \overline{\operatorname{Nm}_{K/F}(\underline{K}^\times)}$, so $a \in \operatorname{Nm}_{K_{w'}/F_{v'}}(K_{w'}^\times)$ for all $w' \mid v'$ with $v' \neq v$, and $a \notin \operatorname{Nm}_{K_v/F_v}(K_v^\times)$; but then a is a local norm at all but one real place, so by the Hasse norm theorem (Theorem 26.8.23), $a \in \operatorname{Nm}_{K/F}(K^\times)$ is a global norm, but this contradicts that a is not a local norm at v.

Now let $\underline{p} = \prod_{v\in\Sigma} \underline{\pi_v}$. Since $\#\Sigma$ is even, by the previous paragraph and (27.5.16) we get

$$\underline{p} = b\underline{u} \in F^\times \operatorname{Nm}_{K/F}(\underline{K}^\times).$$

Consider the quaternion algebra $\left(\dfrac{K, b}{F}\right)$. For all places $v \in \Sigma$, we have K_v a field and $b = \pi_v u_v^{-1} \notin \operatorname{Nm}_{K_v/F_v}(K_v^\times)$, so $v \in \operatorname{Ram} B$. At every other place $v' \notin \Sigma$, we have either that K_v is not a field or $b = u_v^{-1} \in \operatorname{Nm}_{K_v/F_v}(K_v^\times)$, and in either case $v' \notin \operatorname{Ram} B$. $\square$

27.6 Noncommutative adeles

We retain notation from the previous section, in particular the abbreviations $R = R_{(S)}$ and $\underline{F}_{\$} = \widehat{F}$. Let B be a finite-dimensional simple algebra over the global field F. In this section, we extend idelic notions to B; the main case of interest is where B is a quaternion algebra over F. We recall the topology on B_v for places v, discussed in section 13.5. Let $O \subseteq B$ be an R-order.

27.6.1. The **adele ring** of B is the restricted direct product of the topological rings B_v with respect to O_v:

$$\underline{B} := \prod_{v\in\operatorname{Pl}(F)}{}' B_v = \left\{(\alpha_v)_v \in \textstyle\prod_v B_v : \alpha_v \in O_v \text{ for all but finitely many } v \notin S\right\}$$

The topology on $\underline{B}$ (as a topological ring) is uniquely characterized by the property that the subring $\prod_v O_v$ is open with the product topology.

By the local-global dictionary for lattices (Theorem 9.4.9), the definition of $\underline{B}$ is independent of the choice of order O and eligible set S (and base ring $R = R_{(S)}$).

27.6.2. Just as in 27.4.5, we embed $B \hookrightarrow \underline{B}$ diagonally. A basis for B as an F-vector space shows that $\underline{B}$ is a free $\underline{F}$-module of finite rank. Then, since the image $F \hookrightarrow \underline{F}$ is discrete, closed, and cocompact, arguing in each coordinate (with respect to the chosen basis), we conclude that $B \hookrightarrow \underline{B}$ is discrete, closed, and cocompact. (Details are requested in Exercise 27.11.)

We now turn to the multiplicative structure, the main object of our concern.

27.6.3. The **idele group** of B is the restricted direct product of the topological groups $B_v^\times$ with respect to $O_v^\times$:

$$\underline{B}^\times := \prod_v{}' B_v^\times = \left\{(\alpha_v)_v \in \textstyle\prod_v B_v^\times : \alpha_v \in O_v^\times \text{ for all but finitely many } v\right\};$$

equivalently, $\underline{B}^\times$ is the unit group of $\underline{B}$ with the topology as in 27.4.10. The topology on $\underline{B}^\times$ as a topological group is characterized by the condition that the subgroup $\prod_v O_v^\times$

is open with the product topology. Again, $\underline{B}^\times$ is independent of the choice of O and eligible set S because any two such constructions differ at only finitely many places.

27.6.4. The ***S*-finite adele ring** is

$$\underline{B}_{\$} := \prod_{v \notin S}{}' B_v; \tag{27.6.5}$$

extending by zero, we may identify $\underline{B}_{\$}$ with its image in $\underline{B}$. The S-finite adele ring has a compact open subring

$$\underline{O}_{\$} := \prod_{v \notin S} O_v \subseteq \underline{B}_{\$}. \tag{27.6.6}$$

We similarly define the ***S*-finite idele group** with its compact open subgroup

$$\underline{B}_{\$}^\times := \prod_{v \notin S}{}' B_v^\times \supset \prod_{v \notin S} O_v^\times =: \underline{O}_{\$}^\times. \tag{27.6.7}$$

When no confusion can result (S is clear from context), we will drop the superscript and replace with hats, writing simply $\widehat{B} = \underline{B}_{\$}$ and $\underline{O}_{\$}^\times = \widehat{O}^\times$, etc.

Just as in 27.5.1, the ideles provide a convenient way of encoding fractional ideals, as follows.

Lemma 27.6.8. *The set of locally principal, right fractional O-ideals is in bijection with $\widehat{B}^\times/\widehat{O}^\times$ via the map $I \mapsto \widehat{\alpha}\widehat{O}^\times$, where $I_{\mathfrak{p}} = \alpha_{\mathfrak{p}} O_{\mathfrak{p}}$ and $\widehat{\alpha} = (\alpha_{\mathfrak{p}})_{\mathfrak{p}}$; this map induces a bijection*

$$\begin{aligned} \operatorname{Cls}_{\mathsf{R}} O &\leftrightarrow B^\times\backslash\widehat{B}^\times/\widehat{O}^\times \\ [I]_{\mathsf{R}} &\mapsto B^\times\widehat{\alpha}\widehat{O}^\times. \end{aligned} \tag{27.6.9}$$

Proof. Let I be a locally principal right fractional O-ideal, so $I_{\mathfrak{p}} = \alpha_{\mathfrak{p}} O_{\mathfrak{p}}$ for all primes $\mathfrak{p}$ of R, with $\alpha_{\mathfrak{p}}$ well-defined up to right multiplication by an element of $O_{\mathfrak{p}}^\times$, so to I we associate $(\alpha_{\mathfrak{p}} O_{\mathfrak{p}}^\times)_{\mathfrak{p}} = \widehat{\alpha}\widehat{O}^\times \in \widehat{B}^\times/\widehat{O}^\times$. Conversely, given $\widehat{\alpha} \in \widehat{B}^\times/\widehat{O}^\times$ we recover $I = \widehat{\alpha}\widehat{O} \cap B$ from Lemmas 9.4.6 and 9.5.3.

The equivalence relation defining the (right) class set is given by left multiplication by $B^\times$, so the second statement follows. □

Remark 27.6.10. We recall by Main Theorem 16.6.1 that for B a quaternion algebra, an R-lattice $I \subset B$ is locally principal if and only if it is invertible.

27.6.11. In analogy with 27.4.13, we have a natural multiplicative map

$$\begin{aligned} \|\ \|\colon \underline{B}^\times &\to \mathbb{R}_{>0} \\ \underline{\alpha} = (\alpha_v)_v &\mapsto \prod_v |\operatorname{nrd}(\alpha_v)|_v \end{aligned} \tag{27.6.12}$$

and we define

$$\underline{B}^{(1)} = \ker\|\ \| = \{\underline{\alpha} = (\alpha_v)_v : \textstyle\prod_v |\operatorname{nrd}(\alpha_v)|_v = 1\} \tag{27.6.13}$$

By the product formula (14.4.6), we have $B^\times \leq \underline{B}^{(1)}$.

By comparison, we have also the groups

$$B^1 = \{\alpha \in B : \operatorname{nrd}(\alpha) = 1\}$$
$$\underline{B}^1 = \{\alpha \in \underline{B} : \operatorname{nrd}(\alpha) = 1\}$$

satisfying $B^1 \leq \underline{B}^1 \leq \underline{B}^{(1)}$.

The following theorem is fundamental (see Fujisaki [Fuj58, Theorem 8.3], Weil [Weil82, Lemma 3.1.1]).

Main Theorem 27.6.14 (Fujisaki's lemma). *Let B be a finite-dimensional* division *algebra over a global field F. Then the following statements hold.*

(a) $B^\times \leq \underline{B}^{(1)}$ *is cocompact.*

(b) *For any eligible set S, the subgroup $B^\times \leq \underline{B}^\times_{\$} = \widehat{B}^\times$ is cocompact and the set $B^\times\backslash\widehat{B}^\times/\widehat{O}^\times$ is finite.*

Proof. The natural place to prove this result is after some more serious analysis has been done—but it is too important to wait for this. The small amount of input needed, which can be seen as an (ineffective) idelic version of the Minkowski convex body theorem (Theorem 17.5.5), is as follows. There exists a compact subset $\underline{E}$ such that for all $\underline{\beta} \in \underline{B}^{(1)}$,

$$\text{the map } \underline{B} \to B\backslash\underline{B} \text{ is } \textit{not} \text{ injective when restricted to } \underline{\beta}\underline{E}. \tag{27.6.15}$$

For the proof of (27.6.15), see Exercise 29.11: in a nutshell, there is a measure $\underline{\mu}$ on $\underline{B}$ in which $\underline{\mu}(B\backslash\underline{B}) < \infty$, and a compact $\underline{E}$ with $\underline{\mu}(\underline{E})$ satisfies 27.6.15, as $\underline{\mu}(\underline{\beta}\underline{E}) = \underline{\mu}(\underline{E})$.

We first quickly prove part (b), assuming part (a). We have $\widehat{O}^\times$ open in $\widehat{B}^\times$, so the open cover $\{B^\times\widehat{\alpha}\widehat{O}^\times\}_{\widehat{\alpha}\in\widehat{B}^\times}$ can be reduced to a finite cover, whence the double coset space is finite.

We now turn to prove (a), which we do in steps.

Step 1: Setup. From the set $\underline{E}$ granted above, we define

$$\underline{X} := \underline{E} - \underline{E} = \{\underline{\eta} - \underline{\eta}' : \underline{\eta}, \underline{\eta}' \in \underline{E}\}. \tag{27.6.16}$$

Since $\underline{E}$ is compact, we conclude that $\underline{X} \subseteq \underline{B}$ is compact (subtraction is continuous). We will also use the set of products $\underline{X}\underline{X}$, which is again compact in $\underline{B}$.

Step 2: Measuring differences. We claim that for all $\underline{\beta} \in \underline{B}^{(1)}$, we have $\underline{\beta}\underline{X} \cap B^\times \neq \emptyset$. Indeed, let $\underline{\beta} \in \underline{B}^{(1)}$. By (27.6.15), there exist distinct $\underline{\eta}, \underline{\eta}' \in \underline{E}$ (depending on $\underline{\beta}$) such that

$$\underline{\beta}(\underline{\eta} - \underline{\eta}') \in B. \tag{27.6.17}$$

Because $\underline{\eta} \neq \underline{\eta}'$ and B is a division algebra (here is where we use this hypothesis), we have $\underline{\beta}(\underline{\eta} - \underline{\eta}') \in B^\times$, as claimed.

A similar statement holds on the right, multiplying by $\underline{\beta}^{-1}$.

Step 3: Cocompactness. We claim that there exists a compact set $\underline{K} \subseteq \underline{B}^\times \times \underline{B}^\times$ such that for all $\underline{\beta} \in \underline{B}^{(1)}$, there exist $\beta \in B^\times$ and $\underline{\nu} \in \underline{B}^{(1)}$ such that $\underline{\beta} = \beta\underline{\nu}$ and $(\underline{\nu}, \underline{\nu}^{-1}) \in \underline{K}$.

To prove this claim, first define $T := B^\times \cap \underline{X}\underline{X}$. Since T is the intersection of a discrete set and a compact set, we conclude that T is finite. Let $T^{-1} := \{\gamma^{-1} : \gamma \in T\}$ and let

$$\underline{K} := T^{-1}\underline{X} \times \underline{X}. \tag{27.6.18}$$

To check the claim, let $\underline{\beta} \in \underline{B}^{(1)}$. By Step 2, $\underline{\beta}\underline{X} \cap B^\times \neq \emptyset$ and (similarly) $\underline{X}\underline{\beta}^{-1} \cap B^\times \neq \emptyset$. Therefore there exist $\underline{\nu}, \underline{\nu}' \in \underline{X}$ and $\beta, \beta' \in B^\times$ such that

$$\underline{\beta}\underline{\nu} = \beta \quad \text{and} \quad \underline{\nu}'\underline{\beta}^{-1} = \beta'. \tag{27.6.19}$$

Therefore

$$\beta'\beta = (\underline{\nu}'\underline{\beta}^{-1})(\underline{\beta}\underline{\nu}) = \underline{\nu}'\underline{\nu} \in B^\times \cap \underline{X}\underline{X}.$$

Then $\underline{\nu}^{-1} \in T^{-1}\underline{X}$, and $T^{-1}\underline{X}$ is compact. We have shown that $\underline{\beta} = \beta\underline{\nu}^{-1}$ with $\beta \in B^\times$ and $(\underline{\nu}^{-1}, \underline{\nu}) \in \underline{K} = T^{-1}\underline{X} \times \underline{X}$, and this proves the claim.

Step 4: Conclusion. By the definition of the topology on $\underline{B}^\times$, the set

$$\{\underline{\nu} \in \underline{B}^{(1)} : (\underline{\nu}, \underline{\nu}^{-1}) \in \underline{K}\}$$

is compact; then, by the claim in Step 3, this set surjects onto $B^\times\backslash\underline{B}^{(1)}$. We conclude that $B^\times\backslash\underline{B}^{(1)}$ is compact, completing the proof of part (a). $\square$

Corollary 27.6.20. *If B is a division algebra or $B \simeq \mathrm{M}_n(F)$, and $O \subseteq B$ is an R-order, then the class set $\mathrm{Cls}_\mathsf{R}\, O$ is finite.*

Proof. If B is a division algebra, we combine Lemma 27.6.8 and Main Theorem 27.6.14. Otherwise, $B \simeq \mathrm{M}_n(F)$. If O is maximal then $O \simeq \mathrm{M}_n(R)$ and the result follows from 17.3.7; and then the result for a general order O follows from Exercise 17.3(b). $\square$

Remark 27.6.21. Corollary 27.6.20 covers all quaternion algebras B. This finiteness statement generalizes to the theorem of Jordan–Zassenhaus: see Remark 17.7.27.

27.6.22. The idelic point of view (Lemma 27.6.8) also makes it clear why the class number is independent of the order within its genus (Lemma 17.4.11): the idelic description only depends on the local orders, up to isomorphism.

We have two other objects that admit a nice idelic description.

27.6.23. The genus of an order and its type set (see section 17.4) can be similarly described. Let O be an R-order, and let $O' \in \operatorname{Gen} O$ be an order in the genus of O. Then O' is locally isomorphic to O, so there exists $\widehat{\nu} \in \widehat{B}^\times$ such that $\widehat{\nu}\widehat{O}\widehat{\nu}^{-1} = \widehat{O}'$, well defined up to right multiplication by an element of the normalizer $N_{\widehat{B}^\times}(\widehat{O})$. We recover $O' = \widehat{O}' \cap B$, so this gives a bijection

$$\operatorname{Gen} O \leftrightarrow \widehat{B}^\times / N_{\widehat{B}^\times}(\widehat{O}).$$

Two such orders are isomorphic if and only if there exists $\beta \in B^\times$ such that $\beta O \beta^{-1} = O'$, so we have a bijection

$$\operatorname{Typ} O \leftrightarrow B^\times \backslash \widehat{B}^\times / N_{\widehat{B}^\times}(\widehat{O}). \tag{27.6.24}$$

Corollary 27.6.25. *The type set* Typ O *is finite.*

Proof. In view of (27.6.24), the double coset $B^\times \backslash \widehat{B}^\times / N_{\widehat{B}^\times}(\widehat{O})$ is a further quotient of the set $B^\times \backslash \widehat{B}^\times / \widehat{O}^\times$ which is finite by Main Theorem 27.6.14. □

27.6.26. Referring to section 18.5 locally, we see that the group of locally principal two-sided O-ideals $\operatorname{Idl}(O)$ is in bijection with

$$\widehat{O}^\times \backslash N_{\widehat{B}^\times}(\widehat{O}) / \widehat{O}^\times = N_{\widehat{B}^\times}(\widehat{O}) / \widehat{O}^\times = \widehat{O}^\times \backslash N_{\widehat{B}^\times}(\widehat{O}) \tag{27.6.27}$$

where

$$N_{\widehat{B}^\times}(\widehat{O}) = \{\widehat{\alpha} \in \widehat{B}^\times : \widehat{\alpha}\widehat{O} = \widehat{O}\widehat{\alpha}\}$$

is the normalizer of $\widehat{O}$ in $\widehat{B}^\times$. Furthermore, the group of isomorphism classes of locally principal two-sided O-ideals is therefore in bijection with

$$N_{B^\times}(O) \backslash N_{\widehat{B}^\times}(\widehat{O}) / \widehat{O}^\times = N_{\widehat{B}^\times}(\widehat{O}) / \bigl(N_{B^\times}(O)\widehat{O}^\times\bigr) = \bigl(N_{B^\times}(O)\widehat{O}^\times\bigr) \backslash N_{\widehat{B}^\times}(\widehat{O}).$$

27.7 Reduced norms

To conclude, we consider reduced norms; we specialize and suppose that B is a quaternion algebra.

27.7.1. Since S contains all archimedean places, by the local norm calculation (Lemma 13.4.9), we have $\operatorname{nrd}(\widehat{B}^\times) = \widehat{F}^\times$. By the Hasse–Schilling theorem on norms (Main Theorem 14.7.4), we have $\operatorname{nrd}(B^\times) = F^\times_{>_\Omega 0}$, where $\Omega \subseteq \operatorname{Ram} B$ is the set of real ramified places and $F^\times_{>_\Omega 0}$ is the set of elements positive at all $v \in \Omega$ (recalling 14.7.2). Therefore, the reduced norm (in each component) yields a surjective map

$$\operatorname{nrd} \colon B^\times \backslash \widehat{B}^\times / \widehat{O}^\times \to F^\times_{>_\Omega 0} \backslash \widehat{F}^\times / \operatorname{nrd}(\widehat{O}^\times). \tag{27.7.2}$$

We will now see that the group $F^\times_{>_\Omega 0} \backslash \widehat{F}^\times / \operatorname{nrd}(\widehat{O}^\times)$ in (27.7.2) is a certain class group of R.

Lemma 27.7.3. *The subgroup* $\operatorname{nrd}(\widehat{O}^\times) F^\times_{>_\Omega 0} \leq \widehat{F}^\times$ *is a finite-index open subgroup containing* $F^\times_{>0}$*, the group of totally positive elements of* $F^\times$*. If moreover* O *is maximal, then* $\operatorname{nrd}(\widehat{O}^\times) = \widehat{R}^\times$.

Proof. In Lemma 13.4.9, we saw that if $O_\mathfrak{p}$ is maximal, then $\operatorname{nrd}(O_\mathfrak{p}^\times) = R_\mathfrak{p}^\times$; for the finitely many remaining $\mathfrak{p} \subseteq R$, the $R_\mathfrak{p}$-order $O_\mathfrak{p}$ is of finite index in a maximal superorder, so $\operatorname{nrd}(O_\mathfrak{p}^\times)$ is a finite index open subgroup of $R_\mathfrak{p}^\times$. Putting these together, we conclude $\operatorname{nrd}(\widehat{O}^\times)$ is a finite index open subgroup of $\widehat{R}^\times$.

But $\widehat{F}^\times/\widehat{R}^\times F^\times_{>_\Omega 0} \simeq \mathrm{Cl}_\Omega R$ is a finite group and therefore

$$[\widehat{F}^\times : \mathrm{nrd}(\widehat{O}^\times)F^\times_{>_\Omega 0}] = [\widehat{F}^\times : \widehat{R}^\times F^\times_{>_\Omega 0}][\widehat{R}^\times F^\times_{>_\Omega 0} : \mathrm{nrd}(\widehat{O}^\times)F^\times_{>_\Omega 0}] < \infty.$$

Finally, we have $F^\times_{>_\Omega 0} \geq F^\times_{>0}$, as the latter possibly requires further positivity. □

27.7.4. Let

$$G(O) := F^\times_{>_\Omega 0}\,\mathrm{nrd}(\widehat{O}^\times) \leq \widehat{F}^\times. \tag{27.7.5}$$

From Lemma 27.7.3, $G(O)$ is a finite-index open subgroup containing $F^\times_{>0}$. By class field theory 27.5.12, there exists a class field K for $G(O)$, i.e., there exists a finite abelian extension $K \supseteq F$ and an Artin isomorphism

$$\mathrm{Cl}_{G(O)} R = F^\times/G(O) \xrightarrow{\sim} \mathrm{Gal}(K \,|\, F). \tag{27.7.6}$$

The group $\mathrm{Cl}_{G(O)} R$ only depends on the *genus* of O: if $O' \in \mathrm{Gen}\, O$ then O' is locally isomorphic to O, so there exists $\widehat{\nu} \in \widehat{B}^\times$ such that $\widehat{O}' = \widehat{\nu}^{-1}\widehat{O}\widehat{\nu}$ so $\mathrm{nrd}(\widehat{O}'^\times) = \mathrm{nrd}(\widehat{O}^\times)$.

Example 27.7.7. Suppose F is a number field and S is the set of archimedean places of F, so that $R = \mathbb{Z}_F$ is the ring of integers in F. Suppose further that O is maximal. Then $G(O) = F^\times_{>_\Omega 0}\widehat{R}^\times$. Recalling 17.8.2, let Ω be the set of ramified (necessarily real) archimedean places of B, and let $\mathrm{Cl}_\Omega \mathbb{Z}_F := F^\times_{>_\Omega 0}\backslash \widehat{F}^\times/\widehat{R}^\times$, equivalently, $\mathrm{Cl}_\Omega \mathbb{Z}_F$ is the group of fractional ideals of $\mathbb{Z}_F$ modulo the subgroup of principal fractional ideals generated by elements in $F^\times_{>_\Omega 0}$. Then $\mathrm{Cl}_{G(O)} \mathbb{Z}_F = \mathrm{Cl}_\Omega \mathbb{Z}_F$ by definition, so we have a surjective map

$$\mathrm{nrd}\colon\ \mathrm{Cls}\, O \to \mathrm{Cl}_{G(O)} \mathbb{Z}_F = \mathrm{Cl}_\Omega \mathbb{Z}_F.$$

The two extreme cases: if B is unramified at all real places, then $\Omega = \emptyset$, and $\mathrm{Cl}_\Omega \mathbb{Z}_F = \mathrm{Cl}\, \mathbb{Z}_F$ is the class group; if B is ramified at all real places, then $\mathrm{Cl}_\Omega \mathbb{Z}_F = \mathrm{Cl}^+ \mathbb{Z}_F$ is the narrow class group.

Remark 27.7.8. It is a fundamental result of Eichler (Theorem 17.8.3) that whenever there exists $v \in S$ such that B is unramified at v, then the reduced norm map (27.7.2) is injective, and hence *bijective*, giving a bijection between the class set of O and a certain class group of R. This topic is the main result in the next chapter.

Exercises

Unless otherwise specified, let F be a global field and let B be a quaternion algebra over F.

1. If we take the direct product instead of the restricted direct product in the definition of the adele ring, we lose local compactness. More precisely, let $\{X_i\}_{i\in I}$ be a collection of nonempty topological spaces. Show that $X = \prod_{i\in I} X_i$ is locally compact if and only if each X_i is locally compact and all but finitely many X_i are compact.
2. Review the language of group actions and fundamental sets (section 34.1).

a) Equip $\mathbb{Q}$ with the discrete topology. We have a group action $\mathbb{Q} \circlearrowright \underline{\mathbb{Q}}$ (by addition). Show that $\widehat{\mathbb{Z}} \times [0,1] \subseteq \widehat{\mathbb{Q}} \times \mathbb{R}$ is a fundamental set for this action. *[Hint: Review the arguments in Lemmas 27.1.13–27.1.14.]*
b) Similarly, show that $\mathbb{Q}^\times \circlearrowright \underline{\mathbb{Q}}^\times$ and that $\widehat{\mathbb{Z}}^\times \times \mathbb{R}_{>0} \subseteq \widehat{\mathbb{Q}} \times \mathbb{R}$ is a fundamental set for this action.

3. For a prime p, let $\widehat{p} = (p, 1, \dots, 1, p, 1, \dots) \in \underline{\mathbb{Q}}$ be the adele which is equal to p in the pth and ∞th component and 1 elsewhere.

a) Show that the sequence $\widehat{p}$, ranging over primes p, does not converge in $\underline{\mathbb{Q}}^\times$; conclude that $\underline{\mathbb{Q}}^\times$ is not compact.
b) However, show that the sequence $\widehat{p}$ has a subsequence converging to the identity in the quotient $\underline{\mathbb{Q}}^\times/\mathbb{Q}^\times$.

4. Recall that $\widehat{\mathbb{Z}} = \varprojlim_n \mathbb{Z}/n\mathbb{Z} \simeq \prod_p \mathbb{Z}_p$.

(a) Prove that each $\widehat{\alpha} \in \widehat{\mathbb{Z}}$ has a unique representation as $\widehat{\alpha} = \sum_{n=1}^{\infty} c_n n!$ where $c_n \in \mathbb{Z}$ and $0 \le c_n \le n$.
(b) Prove that $\widehat{\mathbb{Z}}^\times \simeq \widehat{\mathbb{Z}} \times \prod_{n=1}^{\infty} \mathbb{Z}/n\mathbb{Z}$ as profinite groups. *[Hint: Consider the product of the p-adic logarithm maps and use the fact that for every prime power p^e there are infinitely many primes q such that $p^e \parallel (q-1)$.]*
(c) Prove for every $n \in \mathbb{Z}_{>0}$ that the natural map $\mathbb{Z}/n\mathbb{Z} \to \widehat{\mathbb{Z}}/n\widehat{\mathbb{Z}}$ is an isomorphism.
(d) Prove that there is a bijection from $\mathbb{Z}_{\ge 0}$ to the set of open subgroups of $\widehat{\mathbb{Z}}$ mapping $n \mapsto n\widehat{\mathbb{Z}}$.

5. We recall from 27.1.15 the *solenoid* $\mathrm{Sol} = (\mathbb{R} \times \widehat{\mathbb{Z}})/\mathbb{Z}$ (with $\mathbb{Z}$ embedded diagonally, and given the quotient topology).

(a) Prove that Sol is a compact, Hausdorff, and connected topological group.
(b) Prove that $\mathrm{Sol} \simeq \underline{\mathbb{Q}}/\mathbb{Q}$ as topological groups.
(c) Prove that $\mathrm{Sol} \simeq \varprojlim_n \mathbb{R}/\frac{1}{n}\mathbb{Z}$ with respect to the directed system $n \in \mathbb{Z}_{\ge 1}$ under divisibility.
(d) Show that the group of path components of Sol is isomorphic to $\widehat{\mathbb{Z}}/\mathbb{Z}$, and conclude that Sol is not path connected. *[Hint: the neutral path component is the image of $\{0\} \times \mathbb{R} \subseteq \widehat{\mathbb{Q}} \times \mathbb{R}$.]*

6. Recall that by definition, a set $U \subseteq \underline{F}$ is open if and only if $U \cap \underline{F}_S$ is open in $\underline{F}_S$ for all eligible $S \subseteq \mathrm{Pl}\, F$. Show that $U \subseteq \underline{F}$ is open if and only if for all $\underline{a} \in \underline{F}$ that $(\underline{a} + U) \cap \prod_v R_v$ is open in $\prod_v R_v$.
7. Show that the topology on $\underline{F}^\times$ agrees with the subspace topology induced on $\underline{F}^\times \hookrightarrow \underline{F} \times \underline{F}$ by $x \mapsto (x, x^{-1})$.

8. Show that if S is eligible and $O \subseteq B$ is an $R_{(S)}$-order, then
$$\underline{B} = \left\{(x_v)_v \in \textstyle\prod_v B_v : x_v \in O_v \text{ for all but finitely many } v\right\}$$
and
$$\underline{B}^\times = \left\{(x_v)_v \in \textstyle\prod_v B_v^\times : x_v \in O_v^\times \text{ for all but finitely many } v\right\}$$
and therefore that this definition is independent of the choice of O (and S).

▹ 9. Prove Lemma 27.5.6.

10. Returning to 27.5.13, let $D_F \leq C_F$ be the connected component of 1 in the idele class group of F. Show that D_F is a closed subgroup with
$$D_F \simeq \mathbb{R} \times (\mathbb{R}/\mathbb{Z})^c \times \mathrm{Sol}^{r+c-1}$$
where r is the number of real places of F and c the number of complex places. Interpret this isomorphism explicitly for F a quadratic field for both F real and imaginary: what 'explains' the factors that appear?

▹ 11. Let B be a finite-dimensional F-algebra.

(a) Show that B is discrete and closed in $\underline{B}$. *[Hint: F is discrete in $\underline{F}$ by the product formula.]*
(b) Show that B is cocompact in $\underline{B}$ (under the diagonal embedding), i.e., that $\underline{B}/B$ is compact.

12. Give a fundamental system of neighborhoods of 0 in $\widehat{B}$ and of 1 in $\widehat{B}^\times$.

▹ 13. Let A be a topological ring.

(a) Suppose that $A^\times \subseteq A$ has the induced topology. Give an example to show that the map $x \mapsto x^{-1}$ on $A^\times$ is not necessarily continuous.

(b) Now embed
$$\begin{aligned} A^\times &\hookrightarrow A \times A \\ x &\mapsto (x, x^{-1}) \end{aligned}$$
and give $A^\times$ the subspace topology. Show that $A^\times$ in this topology is a topological group.

▹ 14. Let $S \subset \mathrm{Pl}\, F$ be an eligible set.

(a) Show that $R_{(S)}$ is discrete in $F_S = \prod_{v\in S} F_v$. *[Hint: it is enough to show this for a neighborhood of* 0, *and then use the fact that the norm must be an integer.]*
(b) Prove that if O is an $R_{(S)}$-order in B, then O is discrete in $B_S = \prod_{v\in S} B_v$.

15. Let F be a global field and let K be a finite separable extension of F.

(a) Show that $\underline{K} \simeq \underline{F} \otimes_F K$. *[Hint: Use the fact that $F_v \otimes_F K \simeq \prod_w K_w$ where w runs over the places above v.]*

(b) Show that

$$\underline{K} = \left\{(x_w)_w \in \prod_w K_w : |\mathrm{Nm}_{K_w/F_v}\, x_v|_v \leq 1 \text{ for almost all } v\right\}$$

but that the inclusion

$$\underline{B} = B \otimes_F \underline{F} \subset \left\{(x_v)_v \in \prod_v B_v : |\mathrm{nrd}(x_v)|_v \leq 1 \text{ for almost all } v\right\}$$

is strict, so the corresponding statement is *false* for B.

16. Let $R = R_{(S)}$ be a global ring and O be an R-order in B. Show that the set of R-orders which are connected to O is in bijection with $\widehat{B}^\times / N_{\widehat{B}^\times}(\widehat{O})$, where $N_{\widehat{B}^\times}(\widehat{O})$ is the normalizer of $\widehat{O}$ in $\widehat{B}^\times$.

17. Extend Lemma 27.7.3 as follows: if O is an Eichler order, then $\mathrm{nrd}(\widehat{O}^\times) = \widehat{R}^\times$.

This chapter is licensed under the terms of the Creative Commons Attribution-NonCommercial 4.0 International License (http://creativecommons.org/licenses/by-nc/4.0/), which permits any noncommercial use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license and indicate if changes were made.

The images or other third party material in this chapter are included in the chapter's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the chapter's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.

Chapter 28

Strong approximation

28.1 ▹ Beginnings

We have already seen in several places in this book how theorems about quaternion algebras over global fields are often first investigated locally, and then a global result is recovered using some form of approximation. Approximation provides a way to transfer analytic properties (encoded in congruences or bounds) into global elements. In this chapter, we develop robust approximation theorems and investigate their arithmetic applications.

We begin by reviewing weak and strong approximation over $\mathbb{Q}$, taking a breath in preparation for the idelic efforts to come.

28.1.1. The starting point is the Sun Zi theorem (CRT): given a finite, nonempty set S of primes, and for each $p \in S$ an exponent $n_p \in \mathbb{Z}_{\geq 1}$ and an element $x_p \in \mathbb{Z}/p^{n_p}\mathbb{Z}$, there exists $x \in \mathbb{Z}$ such that $x \equiv x_p \pmod{p^{n_p}}$ for all $p \in S$. These congruences can be equivalently formulated in the p-adic metric by lifting to $x_p \in \mathbb{Z}_p$ and asking for $|x - x_p| < p^{-n_p}$ for $p \in S$; or equivalently, the map $\mathbb{Z} \to \prod_{p \in S} \mathbb{Z}_p$ has dense image for any finite, nonempty set of primes S (giving the target the product topology). We may therefore think of the CRT as an *approximation* theorem, in the sense that it allows us to find an integer that simultaneously approximates a finite number of p-adic integers arbitrarily well.

We may generalize 28.1.1 (recalling notation from section 27.1, in particular 27.1.8) as follows.

Theorem 28.1.2 (Weak approximation). *Let $S \subseteq \operatorname{Pl}\mathbb{Q}$ be a finite, nonempty set of places of $\mathbb{Q}$. Then the image of $\mathbb{Q} \hookrightarrow \mathbb{Q}_S := \prod_{v \in S} \mathbb{Q}_v$ is dense.*

Proof. Let $x_v \in \mathbb{Q}_v$ for each $v \in S$, and let $\epsilon > 0$. We want to show

$$\text{there exists } x \in \mathbb{Q} \text{ such that } |x - x_v|_v < \epsilon \text{ for all } v \in S. \tag{28.1.3}$$

We proceed by considering increasingly more general cases; our intent is to communicate (concrete, effective) meaning behind the symbols and to set us up to prove

© The Author(s) 2021

J. Voight, *Quaternion Algebras*, Graduate Texts in Mathematics 288,
https://doi.org/10.1007/978-3-030-56694-4_28

strong approximation below; for a short, uniform proof (which reproves the CRT), see Exercise 28.1.

Case 1. Suppose $\infty \notin S$ and $x_p \in \mathbb{Z}_p$ for all $p \in S$. Then (28.1.3) holds by the CRT, as in 28.1.1; in fact, we have infinitely many $x \in \mathbb{Z}$ for this purpose.

Case 2. Suppose $\infty \notin S$, but $x_p \in \mathbb{Q}_p$ for $p \in S$. We employ continuity of multiplication to reduce to the previous case, as follows. We consider the least common denominator:

$$d := \prod_{p \in S} p^{\max(0, -v_p(x_p))} \in \mathbb{Z}_{>0}. \tag{28.1.4}$$

Then $dx_p \in \mathbb{Z}_p$ for all $p \in S$. By the case just established by the CRT, there exists $x' \in \mathbb{Z}$ such that $|x' - dx_p|_p < \epsilon|d|_p$ for all $p \in S$, so taking $x := x'/d$ and dividing through we conclude that (28.1.3) holds.

Case 3. To conclude, suppose $\infty \in S$. We employ an additive translation: we find a rational number close to x_∞ and add to it a small solution to the previous case, as follows. Since $\mathbb{Q} \subseteq \mathbb{R}$ is dense, there exists $y \in \mathbb{Q}$ such that $|y - x_\infty|_\infty < \epsilon/2$. Let $y_p := x_p - y$. From case 2, we find $y' \in \mathbb{Q}$ such that $|y' - y_p|_p < \epsilon$ for all $p \in S$. By case 1, there exist infinitely many $m \in \mathbb{Z}$ such that $|1 - m|_p < \min(1, |y' - y_p|_p)$ for all $p \in S$; for such m, we have $m \equiv 1 \pmod{p}$ so $|m|_p = 1$ and

$$|(y'/m) - y_p|_p = \frac{|y' - my_p|_p}{|m|_p} = |y' - y_p + (1-m)y_p|_p = |y' - y_p|_p < \epsilon \tag{28.1.5}$$

by the ultrametric inequality. By choosing m large enough, we may ensure that $|y'/m|_\infty < \epsilon/2$. Let $x := y'/m + y$. Then by (28.1.5)

$$|x - x_p|_p = |(y'/m) + y - (y_p + y)|_p = |(y'/m) - y_p|_p < \epsilon \tag{28.1.6}$$

and

$$|x - x_\infty|_\infty \leq |x - y|_\infty + |y - x_\infty|_\infty = |y'/m|_\infty + \epsilon/2 < \epsilon \tag{28.1.7}$$

proving (28.1.3). □

Theorem 28.1.8 (Strong approximation). *Let $S \subsetneq \operatorname{Pl}\mathbb{Q}$ be a nonempty set of places. Then the image of $\mathbb{Q} \hookrightarrow \underline{\mathbb{Q}}_{\not S} := \prod'_{v \notin S} \mathbb{Q}_v$ is dense.*

28.1.9. Written out in the standard basis of open sets, strong approximation is equivalent to: given a finite set $T \subseteq \operatorname{Pl}\mathbb{Q}$ disjoint from S, elements $x_v \in \mathbb{Q}_v$ for $v \in T$, and $\epsilon > 0$, there exists $x \in \mathbb{Q}$ such that $|x - x_v|_v < \epsilon$ for all $v \in T$ and $x \in \mathbb{Z}_p$ for all $p \notin S \sqcup T$ with $p \neq \infty$.

Weak approximation follows from strong approximation by forgetting S and weakly approximating x_v for $v \in T$. Indeed, the difference between the 'weak' and the 'strong' is meaningful here. In weak approximation, we satisfy only a finite number of conditions, with no control over the rational number at places $v \notin S$. By contrast, in strong approximation, the role of the set S is switched, and have specified conditions at *all*

primes $v \notin S$, either by approximation at finitely many places in T as in 28.1.9 or by the assertion of integrality at the rest. (This results in some asymmetry in the conclusion for the real place; we could restore this by defining $\mathbb{Z}_\infty := (-1, 1)$, but this would not add content as we could just include this interval in T.)

Proof of Theorem 28.1.8. We prove the statement in its formulation 28.1.9. Naturally, we return to the proof of weak approximation. Without loss of generality (proving a stronger statement), we may assume $\#S = 1$.

If $S = \{\infty\}$, we apply step 2 of weak approximation over the set T: the result $x = x'/d$ already has $x \in \mathbb{Z}_q$ for $q \notin T$, since then $q \nmid d$.

So suppose $S = \{\ell\}$ with ℓ prime. We return to case 3. To define y, we note instead that $\mathbb{Z}[1/p] \subseteq \mathbb{R}$ is dense, so we may take $y \in \mathbb{Z}[1/\ell]$, so in particular $y \in \mathbb{Z}_q$ for $q \neq \ell$. We just showed that we may take $y' \in \mathbb{Z}_q$ for $q \notin T$. And for the integers m, we claim we may take $m = \ell^k$ for $k \in \mathbb{Z}_{\geq 0}$: indeed, we are applying case 1 (CRT) and, as in 28.1.1, this asks for $m \equiv 1 \pmod{p^{n_p}}$ for $p \in T$ (with $n_p \in \mathbb{Z}_{\geq 1}$, and $\ell \neq p$), so we just need to take k to be a common multiple of the orders of $\ell \in (\mathbb{Z}/p^{n_p}\mathbb{Z})^\times$. With this strengthening, we have $x = y'/m + y \in \mathbb{Z}_q$ for all $q \notin S \sqcup T$. □

28.1.10. As is hopefully evident from the proof of strong approximation when $S = \{\infty\}$, aside from continuity of multiplication, the key statement needed was that the map $\mathbb{Z} \to \mathbb{Z}/m\mathbb{Z}$ is surjective for all $m \in \mathbb{Z}$, as provided by the CRT. Or more zippily, what is needed is that the image of $\mathbb{Z} \hookrightarrow \widehat{\mathbb{Z}}$ is dense.

28.1.11. In weak approximation, we can replace the additive group $\mathbb{Q}$ with with the multiplicative group $\mathbb{Q}^\times$: the image of $\mathbb{Q}^\times \hookrightarrow \prod_{v \in S} \mathbb{Q}_v^\times$ is dense *a fortiori*.

However, the embedding $\mathbb{Q}^\times \hookrightarrow \underline{\mathbb{Q}}_{\not S}^\times$ is *not* dense: that is to say, we do not have strong approximation for $\mathbb{Q}^\times$. Indeed, taking $S = \{\infty\}$ we have $\underline{\mathbb{Q}}_{\not S}^\times = \widehat{\mathbb{Q}}^\times$; and since $\widehat{\mathbb{Z}}^\times \cap \mathbb{Q}^\times = \mathbb{Z}^\times = \{\pm 1\}$, the open set $\widehat{\mathbb{Z}}^\times \smallsetminus \{\pm 1\}$ is disjoint from $\mathbb{Q}^\times$. In view of 28.1.10, the problem is also indicated by the fact that $\mathbb{Z}^\times = \{\pm 1\}$ does not surject onto $(\mathbb{Z}/m\mathbb{Z})^\times$ for $m \geq 7$.

28.2 ▹ Strong approximation for $\mathrm{SL}_2(\mathbb{Q})$

We now consider approximation in the noncommutative context. For motivation in this second phase of the introduction, we consider the simplest case where $B = \mathrm{M}_2(\mathbb{Q})$ and take $S = \{\infty\}$, and $\underline{\mathbb{Q}}_{\not S} = \widehat{\mathbb{Q}}$; by analogy, this is like considering a noncommutative generalization of the CRT.

28.2.1. Recall that $B_v = \mathrm{M}_2(\mathbb{Q}_v) \simeq \mathbb{Q}_v^4$ has the coordinate topology (see section 13.5); therefore weak and strong approximation for $B = \mathrm{M}_2(\mathbb{Q})$ follow from these statements for $\mathbb{Q}$, and weak approximation for $\mathrm{GL}_2(\mathbb{Q})$ follows as the determinant is continuous.

28.2.2. We should not expect the embedding $\mathrm{GL}_2(\mathbb{Q}) \hookrightarrow \mathrm{GL}_2(\widehat{\mathbb{Q}})$ to be dense any more than it was for $\mathbb{Q}^\times = \mathrm{GL}_1(\mathbb{Q})$, as in 28.1.11. In fact, we rediscover the same issue by taking determinants: the map $\mathrm{GL}_2(\mathbb{Z}) \to \mathrm{GL}_2(\mathbb{Z}/m\mathbb{Z})$ is not surjective, because $\det(\mathrm{GL}_2(\mathbb{Z})) = \pm 1$ whereas $\det(\mathrm{GL}_2(\mathbb{Z}/m\mathbb{Z})) = (\mathbb{Z}/m\mathbb{Z})^\times$.

Once we restrict to the subgroup of determinant 1, we find a dense subgroup once again.

Theorem 28.2.3. *The image of* $\mathrm{SL}_2(\mathbb{Q}) \hookrightarrow \mathrm{SL}_2(\widehat{\mathbb{Q}})$ *is dense.*

Theorem 28.2.3 is known as **strong approximation** for the group $\mathrm{SL}_2(\mathbb{Q})$. We give a quick proof of Theorem 28.2.3 in two steps. In preparation, we recall from 27.2.6 that

$$\widehat{\mathbb{Q}}^\times = \mathbb{Q}^\times \widehat{\mathbb{Z}}^\times$$

("denominators can be handled globally") and prove an analogous decomposition.

Lemma 28.2.4. *We have*

$$\begin{aligned} \mathrm{GL}_2(\widehat{\mathbb{Q}}) &= \mathrm{GL}_2(\mathbb{Q})\,\mathrm{GL}_2(\widehat{\mathbb{Z}}) \\ \mathrm{SL}_2(\widehat{\mathbb{Q}}) &= \mathrm{SL}_2(\mathbb{Q})\,\mathrm{SL}_2(\widehat{\mathbb{Z}}). \end{aligned} \tag{28.2.5}$$

Proof. We begin with the first statement. The inclusion $\mathrm{GL}_2(\mathbb{Q})\,\mathrm{GL}_2(\widehat{\mathbb{Z}}) \subseteq \mathrm{GL}_2(\widehat{\mathbb{Q}})$ holds; we prove the other containment. Let $\widehat{\alpha} \in \mathrm{GL}_2(\widehat{\mathbb{Q}})$. Consider the collection of lattices $(L_p)_p$ with $L_p = \alpha_p \mathbb{Z}_p^2 \subseteq \mathbb{Q}_p^2$. Since $\alpha_p \in \mathrm{GL}_2(\mathbb{Z}_p)$ for all but finitely many p, we have $L_p = \mathbb{Z}_p^2$ for all but finitely many p. By the local-global dictionary for lattices (Theorem 9.4.9), there exists a unique lattice $L \subseteq \mathbb{Q}^2$ whose completions are L_p. We now rephrase this adelically (and succinctly): letting $\widehat{L} = \widehat{\alpha}\widehat{\mathbb{Z}}^2 \subseteq \widehat{\mathbb{Q}}^2$, we take $L = \widehat{L} \cap \mathbb{Q}^2$. Choose a basis for L and put the columns in a matrix α, so $L = \alpha\mathbb{Z}^2$. Then $\widehat{L} = \alpha\widehat{\mathbb{Z}}^2 = \widehat{\alpha}\widehat{\mathbb{Z}}^2$, and there exists $\gamma \in \mathrm{GL}_2(\widehat{\mathbb{Z}})$ such that $\widehat{\alpha} = \alpha\widehat{\gamma}$. This completes the inclusion.

To get down to SL_2, we take determinants. Let $\widehat{\alpha} \in \mathrm{SL}_2(\widehat{\mathbb{Q}})$ and write it as $\widehat{\alpha} = \alpha\widehat{\gamma}$ with $\alpha \in \mathrm{GL}_2(\mathbb{Q})$ and $\widehat{\gamma} \in \mathrm{GL}_2(\widehat{\mathbb{Z}})$. Then

$$1 = \det(\widehat{\alpha}) = \det(\alpha)\det(\widehat{\gamma}) \in \mathbb{Q}^\times\widehat{\mathbb{Z}}^\times = \widehat{\mathbb{Q}}^\times$$

but $\mathbb{Q}^\times \cap \widehat{\mathbb{Z}}^\times = \{\pm 1\}$; multiplying both $\alpha, \widehat{\gamma}$ by $\begin{pmatrix} -1 & 0 \\ 0 & 1 \end{pmatrix}$ on the right and left respectively, if necessary, we may take $\det(\alpha) = \det(\widehat{\gamma}) = 1$, i.e., $\alpha \in \mathrm{SL}_2(\mathbb{Q})$ and $\widehat{\gamma} \in \mathrm{SL}_2(\widehat{\mathbb{Z}})$. □

Now for the slightly magical second step.

Theorem 28.2.6. *The map*

$$\mathrm{SL}_2(\mathbb{Z}) \to \mathrm{SL}_2(\mathbb{Z}/m\mathbb{Z})$$

is surjective for all $m \in \mathbb{Z}$*; equivalently, the image of* $\mathrm{SL}_2(\mathbb{Z}) \hookrightarrow \mathrm{SL}_2(\widehat{\mathbb{Z}})$ *is dense.*

The statement is nontrivial: a matrix modulo m can certainly be lifted to a matrix in $\mathbb{Z}$ whose determinant will be congruent to 1 modulo m, but the hard part is to ensure that the lifted matrix has determinant equal to 1.

Proof of Theorem 28.2.6. Let $\alpha \in \mathrm{M}_2(\mathbb{Z})$ be such that α maps to the desired matrix in $\mathrm{SL}_2(\mathbb{Z}/m\mathbb{Z})$, so in particular $\det(\alpha) \equiv 1 \pmod m$. By the theory of elementary divisors (Smith normal form), there exist matrices $\mu, \nu \in \mathrm{SL}_2(\mathbb{Z})$ such that $\mu\alpha\nu$ is diagonal; so without loss of generality, we may suppose that $\alpha = \begin{pmatrix} a & 0 \\ 0 & b \end{pmatrix}$ with $ab \equiv 1 \pmod m$. Let

$$\alpha' = \begin{pmatrix} a & -(1-ab) \\ 1-ab & b(2-ab) \end{pmatrix}. \tag{28.2.7}$$

Then $\alpha' \equiv \alpha \pmod m$ and

$$\det(\alpha') = ab(2-ab) + (1-ab)^2 = 1 \tag{28.2.8}$$

so $\alpha' \in \mathrm{SL}_2(\mathbb{Z})$, as claimed. (Compare Shimura [Shi71, Lemma 1.38].) □

Remark 28.2.9. The proof of Theorem 28.2.6 extends in two ways. First, we can replace $\mathbb{Z}$ with a PID and the same proof works. Second, arguing by induction, one can show that the map $\mathrm{SL}_n(\mathbb{Z}) \to \mathrm{SL}_n(\mathbb{Z}/m\mathbb{Z})$ is surjective for all $n \geq 2$ and $m \in \mathbb{Z}$.

We are now ready to prove strong approximation for $\mathrm{SL}_2(\mathbb{Q})$.

Proof of Theorem 28.2.3. Consider the closure of $\mathrm{SL}_2(\mathbb{Q})$ in $\mathrm{SL}_2(\widehat{\mathbb{Q}})$ in the idelic topology; we obtain a closed subgroup. Since $\mathrm{SL}_2(\mathbb{Q}) \geq \mathrm{SL}_2(\mathbb{Z})$, by Theorem 28.2.6 the closure contains $\mathrm{SL}_2(\widehat{\mathbb{Z}})$, but then by Lemma 28.2.4, it contains all of $\mathrm{SL}_2(\mathbb{Q})\,\mathrm{SL}_2(\widehat{\mathbb{Z}}) = \mathrm{SL}_2(\widehat{\mathbb{Q}})$! Therefore $\mathrm{SL}_2(\mathbb{Q}) \leq \mathrm{SL}_2(\widehat{\mathbb{Q}})$ is dense. □

With the preceding context, we are now ready to state a more general formulation of strong approximation for indefinite quaternion algebras over $\mathbb{Q}$. The following theorem is a special case of Main Theorem 28.5.3.

Theorem 28.2.10 (Strong approximation). *Let B be an indefinite quaternion algebra over $\mathbb{Q}$. Then B^1 is dense in $\widehat{B}^1$.*

If B is definite, then O^1 is a finite group () and so cannot be dense in the infinite idelic group $\widehat{B}^1$. This theorem has the following important applications.

Theorem 28.2.11. *Let O be an Eichler order in an indefinite quaternion algebra over $\mathbb{Q}$. Then the following statements hold.*

(a) *Every locally principal right O-ideal is in fact principal, i.e.,* $\#\operatorname{Cls} O = 1$.
(b) *Every order O' locally isomorphic to O is in fact isomorphic to O, i.e.,* $\#\operatorname{Typ} O = 1$.
(c) *For any integer m, the reduction map $O^1 \to (O/mO)^1$ is surjective.*

Proof. Specialize Main Theorem 28.5.3, Corollary 28.5.6, and Corollary 28.5.14, respectively, using 28.5.16. □

28.3 Elementary matrices

Before embarking on our more general idelic quest, we pause to give a second proof of strong approximation for SL_2 using elementary matrices.

28.3.1. Let R be a domain. An **elementary matrix** (or **transvection**) in $\mathrm{SL}_n(R)$ is a matrix which differs from the identity in one off-diagonal entry; such a matrix acts by an elementary row operation (add a multiple of a row to a different row) on the left and by an elementary column operation on the right. For $n = 2$, the elementary matrices are those of the form $\begin{pmatrix} 1 & b \\ 0 & 1 \end{pmatrix}$ or $\begin{pmatrix} 1 & 0 \\ c & 1 \end{pmatrix}$ with $b, c \in F$.

28.3.2. If F is a field, then $\mathrm{SL}_n(F)$ is generated by elementary matrices by the theory of echelon forms (Exercise 28.3).

Lemma 28.3.3. *Let R be a Euclidean domain. Then $\mathrm{SL}_2(R)$ is generated by elementary matrices.*

Proof. The calculation

$$\begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}\begin{pmatrix} 1 & 0 \\ -1 & 1 \end{pmatrix}\begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix} = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix} =: \eta$$

shows that η is in the subgroup of elementary matrices.

We now follow the usual proof of the elementary divisor theorem. Let $\alpha = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in \mathrm{SL}_2(R)$. First, suppose $b = 0$. Then $\det(\alpha) = ad = 1$; adding a times the second row to the first, we may suppose $b = 1$; then multiplying by η on the right we may suppose $a = 1$; elementary row and column operations then give $b = c = 0$, and then $d = 1$. Similarly, if $a = 0$, multiplying by η gives $b = 0$ and we repeat.

So we may suppose $a, b \neq 0$. By the Euclidean algorithm under the norm N, there exists $q, r \in R$ such that $a = bq + r$ and $N(r) < N(b)$. Applying the elementary matrix which adds $-q$ times the second column to the first, we may suppose $a = 0$ or $N(a) < N(b)$. If $a = 0$, we are done by the previous paragraph; otherwise, we multiply on the right by η which swaps columns, and repeat. Because N takes nonnegative integer values, this procedure terminates after finitely many steps. □

Remark 28.3.4. Lemma 28.3.3 holds for general $n \geq 2$, and it follows from the above by induction: see Exercise 28.4.

This theory of elementary matrices has the following striking consequence.

Proposition 28.3.5. *Let R be a Dedekind domain. Then for all ideals $\mathfrak{m} \subseteq R$, the map*

$$\mathrm{SL}_2(R) \to \mathrm{SL}_2(R/\mathfrak{m})$$

is surjective.

Proof. We may suppose $\mathfrak{m}$ is nonzero. Then by the CRT, $R/\mathfrak{m}$ is a finite product of local Artinian principal ideal rings. Therefore $R/\mathfrak{m}$ is Euclidean and by a generalization of Lemma 28.3.3, the group $\mathrm{SL}_2(R/\mathfrak{m})$ is generated by elementary matrices: see Exercise 28.6. Every elementary matrix in $\mathrm{SL}_2(R/\mathfrak{m})$ lifts to an elementary matrix in $\mathrm{SL}_2(R)$, and the statement follows. □

Corollary 28.3.6. *Let F be a global field, and let $R \subseteq F$ be a global ring with eligible set S. Then the image of the map*

$$\mathrm{SL}_2(F) \hookrightarrow \mathrm{SL}_2(\underline{F}_{\not S}) = {\prod_{v \notin S}}' \mathrm{SL}_2(F_v)$$

is dense.

Proof. For brevity, we write $\widehat{F} = \underline{F}_{\not S}$ and $\widehat{R} = \prod_{v \notin S} R_v$. We first show that $\mathrm{SL}_2(R) \hookrightarrow \mathrm{SL}_2(\widehat{R})$ is dense. If $U \subseteq \mathrm{SL}_2(\widehat{R})$ is open, then U contains a standard open neighborhood of the form

$$\{\widehat{\beta} \in \mathrm{SL}_2(\widehat{R}) : \widehat{\beta} \equiv \alpha_{\mathfrak{m}} \pmod{\mathfrak{m}}\}$$

for some $\alpha_{\mathfrak{m}} \in \mathrm{SL}_2(R/\mathfrak{m})$ and $\mathfrak{m} \subseteq R$. The surjectivity in Proposition 28.3.5 then implies that $U \cap \mathrm{SL}_2(R) \neq \emptyset$.

For the statement itself, we again argue with elementary matrices. Let $\widehat{\alpha} = (\alpha_v)_v \in \mathrm{SL}_2(\widehat{F})$; then $\alpha_v \in \mathrm{SL}_2(R_v)$ for all but finitely many v. For these finitely many v, we know $\mathrm{SL}_2(F_v)$ is generated by elementary matrices by Lemma 28.3.2, so by strong approximation in F (Lemma 28.7.2) we can approximate α_v by an element of $\mathrm{SL}_2(F)$ that belongs to any open neighborhood of α_v; for the remaining places we apply the previous paragraph, and we finish using the continuity of multiplication. □

28.4 Strong approximation and the ideal class set

In this section, we provide one more motivation for strong approximation, relating it to the ideal class set as previewed in Eichler's theorem (see section 17.8).

We adopt the following notation for the rest of this chapter. Let R be a global ring with eligible set S and $F = \operatorname{Frac} R$ its global field. Let B be a quaternion algebra over F and let $O \subseteq B$ be an R-order.

28.4.1. By 27.7.1, the reduced norm map

$$\mathrm{nrd}\colon \operatorname{Cls} O = B^\times \backslash \widehat{B}^\times / \widehat{O}^\times \to F^\times_{>_\Omega 0} \backslash \widehat{F}^\times / \mathrm{nrd}(\widehat{O}^\times) \tag{28.4.2}$$

is surjective. Then by class field theory 27.7.4, the codomain $F^\times_{>_\Omega 0} \backslash \widehat{F}^\times / \mathrm{nrd}(\widehat{O}^\times) = \mathrm{Cl}_{G(O)} R$ admits a description as a class group.

The important point that we will soon see: the reduced norm map is injective, therefore bijective, *when strong approximation holds* for the group B^1. But before we get there, we have some explaining to do.

We now investigate the injectivity of the reduced norm map (28.4.2). This map is only a map of (pointed) sets, so first we show that it suffices to look at an appropriate kernel.

28.4.3. For all $\widehat{\beta} \in \widehat{B}^\times$, the map $\widehat{\alpha}O \mapsto \widehat{\alpha}O\widehat{\beta}$ gives a bijection

$$\operatorname{Cls} O = B^\times\backslash\widehat{B}^\times/\widehat{O}^\times \leftrightarrow B^\times\backslash\widehat{B}^\times/\widehat{O}'^\times = \operatorname{Cls} O'$$

where $O' := B \cap \widehat{\beta}^{-1}\widehat{O}\widehat{\beta}$ is connected (locally isomorphic) to O. So it is sensible to consider the maps (28.4.2) for all orders O' connected to O, i.e., the entire genus $\operatorname{Gen} O$. We recall that the type set is finite (Corollary 27.6.25, or Main Theorem 17.7.1 in the number field case using the geometry of numbers).

Our investigations will involve the kernels of the reduced norm maps:

$$B^1 := \{\alpha \in B^\times : \operatorname{nrd}(\alpha) = 1\} \leq \widehat{B}^1 := \{\widehat{\alpha} \in \widehat{B}^\times : \operatorname{nrd}(\widehat{\alpha}) = 1\} \tag{28.4.4}$$

Example 28.4.5. If $B = (a, b \mid F)$, then B^1 admits the Diophantine description

$$B^1 \simeq \{(x, y, z, w) \in F^4 : x^2 - ay^2 - bz^2 + abw^2 = 1\},$$

and the group $\widehat{B}^1$ consists of local solutions at all primes that belong to R_v^4 for almost all places $v \in \operatorname{Pl} F$.

Lemma 28.4.6. *Let $O \subseteq B$ be an R-order. Then the reduced norm map (28.4.2) is injective for all orders $O' \in \operatorname{Gen} O$ if and only if $\widehat{B}^1 \subseteq B^\times\widehat{O}'^\times$ for all $O' \in \operatorname{Gen} O$.*

Proof. If (28.4.2) is injective, then given $\widehat{\alpha} \in \widehat{B}^1$ we have $\operatorname{nrd}(\widehat{\alpha}\widehat{O}^\times) = \operatorname{nrd}(\widehat{O}^\times)$ so $\widehat{\alpha}\widehat{O}^\times = z\widehat{O}^\times$ for some $z \in B^\times$ and $\widehat{\alpha} \in z\widehat{O}^\times \subseteq B^\times\widehat{O}^\times$.

For the converse, since $\operatorname{nrd}\colon B^\times \to F^\times_{>_\Omega 0}$ and $\operatorname{nrd}\colon \widehat{O}^\times \to \operatorname{nrd}(\widehat{O}^\times)$ are both surjective, to show nrd is injective for O we may show that if $\operatorname{nrd}(\widehat{\alpha}) = \operatorname{nrd}(\widehat{\beta}) \in \widehat{F}^\times$ then $\widehat{\alpha}\widehat{O}^\times = z\widehat{\beta}\widehat{O}^\times$ for some $z \in B^\times$. We consider $(\widehat{\alpha}\widehat{\beta}^{-1})(\widehat{\beta}\widehat{O}\widehat{\beta}^{-1}) = (\widehat{\alpha}\widehat{\beta}^{-1})\widehat{O}'$ where as above $O' = B \cap \widehat{\beta}\widehat{O}\widehat{\beta}^{-1} \in \operatorname{Gen} O$. Since $\widehat{\alpha}\widehat{\beta}^{-1} \in \widehat{B}^1$, by hypothesis $\widehat{\alpha}\widehat{\beta}^{-1} = z\widehat{\mu}' = z(\widehat{\beta}\widehat{\mu}\widehat{\beta}^{-1})$ where $z \in B^\times$ and $\widehat{\mu} \in \widehat{O}^\times$, and consequently $\widehat{\alpha}\widehat{O} = z\widehat{\beta}\widehat{\mu}\widehat{O} = z\widehat{\beta}\widehat{O}$, and hence the map is injective. ∎

28.4.7. We have $B^\times\widehat{O}^\times \cap \widehat{B}^1 = B^1\widehat{O}^1$ if and only if $\operatorname{nrd}(O^\times) = F^\times_{>_\Omega 0} \cap \operatorname{nrd}(\widehat{O}^\times)$ (Exercise 28.10).

28.4.8. Suppose that B^1 is dense in $\widehat{B}^1$. Then we claim that $\widehat{B}^1 \subseteq B^1\widehat{O}^1 \subseteq B^\times\widehat{O}^\times$ for all orders $\widehat{O}$. Indeed, if $\widehat{\alpha} = (\alpha_\mathfrak{p})_\mathfrak{p} \in \widehat{B}^1$ then $\widehat{\alpha}\widehat{O}^1 \leq \widehat{B}^1$ is open, so there exists $\alpha \in B^1$ such that $\alpha = \widehat{\alpha}\widehat{\gamma}$ with $\widehat{\gamma} \in \widehat{O}^1$, and $\widehat{\alpha} = \alpha\widehat{\gamma}^{-1} \in B^1\widehat{O}^1$.

We should not expect hypothesis of 28.4.8 to hold for all quaternion algebras: see Exercise 28.7.

28.5 Statement and first applications

In this section, we set up and state the strong approximation theorem, and then derive some applications. Throughout, we abbreviate $\underline{B}_\mathfrak{F} = \widehat{B}$.

Definition 28.5.1. We say B is **S-indefinite** (or B **satisfies the S-Eichler condition**) if S contains a place which is unramified in B.

28.5.2. If F is a number field, then this definition agrees with Definition 17.8.1; and since a complex place is necessarily split and S contains the archimedean places, if B is S-definite over a number field F then F is a totally real number field.

Main Theorem 28.5.3 (Strong approximation). *Let B be a quaternion algebra over a global field and suppose B is S-indefinite. Then B^1 is dense in $\widehat{B}^1$.*

28.5.4. One can think of strong approximation from the following informal perspective: if B_v^1 is *not* compact, then there is enough room for B^1 to "spread out" in B_v^1 so that correspondingly B^1 is dense in the S-finite part $\widehat{B}^1$.

The hypothesis that $B_S^1 = \prod_{v \in S} B_v^1$ is noncompact is certainly necessary for the conclusion that B^1 is dense in $\widehat{B}^1$. Indeed, if $B_S^1 = \prod_{v \in S} B_v^1$ is compact, then since B^1 is discrete in $\underline{B}^1$, the subgroup $B^1 B_S^1 \leq \underline{B}^1$ is closed in $\underline{B}^1$, and $B^1 B_S^1 \neq \underline{B}^1$. On the other hand, if B^1 is dense in $\widehat{B}^1$, then adding the components for $v \in S$ we have $B^1 B_S^1 \leq \underline{B}^1$ dense. This is a contradiction.

We give two proofs of strong approximation over the next two sections. For the moment, we consider some applications.

Our main motivation for strong approximation is the following proposition. We recall the class group 27.7.4 associated to O.

Theorem 28.5.5. *If B is S-indefinite, then the reduced norm map* (28.4.2)

$$\operatorname{nrd}\colon \operatorname{Cls} O = B^\times \backslash \widehat{B}^\times / \widehat{O}^\times \to \operatorname{Cl}_{G(O)} R = F_{>_\Omega 0}^\times \backslash \widehat{F}^\times / \operatorname{nrd}(O^\times)$$

is a bijection for all R-orders $O \subseteq B$: in particular, if I is a locally principal right O-ideal, then I is principal if and only if $\operatorname{nrd}(I)$ *is principal in the class group* $\operatorname{Cl}_{G(O)} R$.

Proof. Combine Lemma 28.4.6 and 28.4.8. ∎

Corollary 28.5.6. *If B is S-indefinite and* $\operatorname{Cl}_{G(O)} R$ *is trivial, then* $\operatorname{Typ} O$ *is trivial, i.e., every order O' locally isomorphic to O is in fact isomorphic to O.*

Proof. The class set $\operatorname{Cls} O$ maps surjectively onto $\operatorname{Typ} O$ by Lemma 17.4.13, and the latter is trivial by Theorem 28.5.5. ∎

28.5.7. More generally, we can grapple with the type set of O, measured by a different (generalized) class group. Recall (27.6.24) that

$$\operatorname{Typ} O \leftrightarrow B^\times \backslash \widehat{B}^\times / N_{\widehat{B}^\times}(\widehat{O}).$$

Let

$$GN(O) := F_{>_\Omega 0}^\times \operatorname{nrd}(N_{\widehat{B}^\times}(\widehat{O})) \leq \widehat{F}^\times. \tag{28.5.8}$$

Since $\widehat{O}^\times \leq N_{\widehat{B}^\times}(\widehat{O})$, we have $GN(O) \geq G(O)$. Define accordingly the class group

$$\operatorname{Cl}_{GN(O)} R = \widehat{F}^\times / GN(O). \tag{28.5.9}$$

Then there is a surjective map of abelian groups

$$\mathrm{Cl}_{G(O)} \to \mathrm{Cl}_{GN(O)} R.$$

Corollary 28.5.10. *If B is S-indefinite, then the reduced norm map induces a bijection*

$$\mathrm{nrd}\colon \mathrm{Typ}\, O \xrightarrow{\sim} \mathrm{Cl}_{GN(O)} R.$$

Proof. We take the further quotient by the normalizer in the bijection in Theorem 28.5.5. □

28.5.11. Returning to 17.4.16, for $B = \mathrm{M}_2(F)$ and $O = \mathrm{M}_2(R)$ we compute that $\mathrm{nrd}(N_{\widehat{B}^\times}(\widehat{O})) = \widehat{F}^2\widehat{R}^\times$, so $\mathrm{Cl}_{GN(O)} R = \mathrm{Cl}\, R/(\mathrm{Cl}\, R)^2$. Thus by Corollary 28.5.10, the types of maximal orders in O are given by $\begin{pmatrix} R & \mathfrak{a} \\ \mathfrak{a}^{-1} & R \end{pmatrix}$ for $[\mathfrak{a}]$ in a set of representatives of $(\mathrm{Cl}\, R)/(\mathrm{Cl}\, R)^2$.

Two other immediate applications of strong approximation that served as motivation are now apparent.

Corollary 28.5.12. *Suppose B is S-indefinite. Then*

$$\widehat{B}^1 = B^1\widehat{O}^1 \quad \textit{and} \quad \underline{B}^1 = B^1\underline{O}^1. \tag{28.5.13}$$

Proof. The inclusion $B^1\widehat{O}^1 \subseteq \widehat{B}^1$ holds, and the converse holds when B^1 is dense in $\widehat{B}^1$ by 28.4.8. For the second statement, we have $\underline{B} = \widehat{B} \times B_S$ and $\underline{O} = \widehat{O} \times B_S$, so we take norm 1 units and multiply both sides of (28.5.13) by $B_S^1 = \prod_{v\in S} B_v^1$. □

Corollary 28.5.14. *Suppose B is S-indefinite. Let $\mathfrak{m} \subseteq R$ be an ideal. Then the reduction map*

$$O^1 \to (O/\mathfrak{m}O)^1$$

is surjective. Moreover, O^1 is dense in $\widehat{O}^1$.

Proof. For all $\alpha_{\mathfrak{m}} \in (O/\mathfrak{m}O)^1$, by strong approximation the open set

$$\{\widehat{\beta} \in \widehat{O}^1 : \widehat{\beta} \equiv \alpha_{\mathfrak{m}} \pmod{\mathfrak{m}O}\} \subseteq \widehat{O}^1$$

contains an element $\alpha \in B^1 \cap \widehat{O}^1 = O^1$ mapping to $\alpha_{\mathfrak{m}}$ in the reduction map. The second statement follows as above from the definition of the topology. □

We now give a name to a large classes of orders where the group $G(O)$ governing principality is explicitly given.

Definition 28.5.15. We say that an R-order $O \subseteq B$ is **locally norm-maximal** if $\mathrm{nrd}(\widehat{O}^\times) = \widehat{R}^\times$.

Equivalently, O is locally norm-maximal if and only if the reduced norm maps $\mathrm{nrd}\colon O_{\mathfrak{p}}^\times \to R_{\mathfrak{p}}^\times$ are surjective for all nonzero primes $\mathfrak{p}$ of R.

Example 28.5.16. If O is maximal, then O is locally norm-maximal (Lemma 13.4.9); more generally if O is Eichler then O is locally norm-maximal (Exercise 23.3).

Certain special cases of Theorem 28.5.5 are important in applications. Recall that $\Omega \subseteq \operatorname{Ram} B$ is the set of real ramified places, and $\operatorname{Cl}_\Omega R$ as defined in 17.8.2 is class group associated to Ω, a quotient of the narrow class group.

Corollary 28.5.17. *Suppose F is a number field and let S be the set of archimedean places of F. Suppose B is S-indefinite and $O \subseteq B$ is locally norm-maximal R-order. Then* $\operatorname{nrd}\colon \operatorname{Cls} O \to \operatorname{Cl}_\Omega R$ *is a bijection.*

Proof. This is just a restatement of Theorem 28.5.5 once we note that $\operatorname{Cl}_{G(O)} R = \operatorname{Cl}_\Omega R$ by Example 27.7.7. □

Proposition 28.5.18. *Let $T \supseteq S$ be a set of primes of R that generate* $\operatorname{Cl}_{G(O)} R$ *and suppose B is T-indefinite. Then every class in* $\operatorname{Cls} O$ *contains an integral (invertible right) O-ideal whose reduced norm is supported in the set T.*

Proof. Let $R_{(T)}$ denote the (further) localization of R at the primes in T. We apply Theorem 28.5.5 to the order $O_{(T)} := O \otimes_R R_{(T)}$: we conclude that there is a bijection $\operatorname{Cls} O_{(T)} \xrightarrow{\sim} \operatorname{Cl}_{G(O_{(T)})} R_{(T)}$. But $\operatorname{Cl}_{G(O_{(T)})} R_{(T)}$ is the quotient of $\operatorname{Cl}_{G(O)} R$ by the primes in T, and so by hypothesis is trivial. Therefore if I is a right O-ideal, then $I_{(T)} := I \otimes_R R_{(T)}$ has $I_{(T)} = \alpha O_{(T)}$ for some $\alpha \in B^\times$. Let $J = \alpha^{-1} I$. Then $[J]_{\mathsf{R}} = [I]_{\mathsf{R}}$ and $J_{\mathfrak{p}} = O_{\mathfrak{p}}$ for all primes $\mathfrak{p} \notin T$ and so J has reduced norm supported in T. Replacing J by aJ with $a \in R$ nonzero and supported in T, we may suppose further that $J \subseteq O$ is integral, and the result follows. □

Example 28.5.19. Let B be a definite quaternion algebra over a totally real (number) field F and let S be the set of archimedean places, so $R = \mathbb{Z}_F$. Let O be a locally norm-maximal R-order in B. Suppose that $\operatorname{Cl}_\Omega R = \{1\}$ and let $\mathfrak{p} \subseteq R$ be a prime of R unramified in B. Then by Proposition 28.5.18, every ideal class in $\operatorname{Cls} O$ contains an integral O-ideal whose reduced norm is a power of $\mathfrak{p}$.

As a special case, we may take $F = \mathbb{Q}$: then $\operatorname{Cl}_\Omega \mathbb{Z} = \operatorname{Cl}^+ \mathbb{Z} = \{1\}$. Therefore, if B is a definite quaternion algebra of discriminant D over $\mathbb{Q}$, and $O \subseteq B$ a locally norm-maximal order (e.g., an Eichler order), then for a prime $p \nmid D$, every invertible right O-ideal class is represented by an integral ideal whose reduced norm is a power of p.

28.6 Further applications

We continue with further applications of strong approximation.

Our next consequence of strong approximation is a refinement the Hasse–Schilling theorem on norms (Main Theorem 14.7.4) as follows.

Theorem 28.6.1 (Eichler's theorem on norms). *Suppose B is S-indefinite, and let $n \in R \cap F^\times_{>_\Omega 0}$. Then there exists $\alpha \in B^\times$* integral over R *such that* $\operatorname{nrd}(\alpha) = n$.

Proof. By Main Theorem 14.7.4, there exists $\alpha \in B^\times$ such that $\operatorname{nrd}(\alpha) = n$. For each prime $\mathfrak{p}$, the set

$$U_\mathfrak{p} = \{\beta_\mathfrak{p} \in B_\mathfrak{p}^1 : \operatorname{trd}(\beta_\mathfrak{p}\alpha) \in R_\mathfrak{p}\} \tag{28.6.2}$$

is (closed and) open since trd is continuous, and further $U_\mathfrak{p}$ is nonempty: if $\mathfrak{p} \in \operatorname{Ram} B$ then already $\alpha_\mathfrak{p}$ is integral over $R_\mathfrak{p}$ and $1 \in U_\mathfrak{p}$, and otherwise $B_\mathfrak{p} \simeq \mathrm{M}_2(F_\mathfrak{p})$ and we may suppose $\alpha_\mathfrak{p} = \begin{pmatrix} 0 & -n \\ 1 & t \end{pmatrix}$ is in rational canonical form whereby

$$\operatorname{trd}\begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}\begin{pmatrix} 0 & -n \\ 1 & t \end{pmatrix} = \operatorname{trd}\begin{pmatrix} 1 & t \\ 0 & n \end{pmatrix} = n + 1 \in R_\mathfrak{p} \tag{28.6.3}$$

shows $\beta_\mathfrak{p} = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix} \in U_\mathfrak{p}$.

Let $\widehat{U} := \left(\prod_\mathfrak{p} U_\mathfrak{p}\right) \cap \widehat{B}^1$; then $\widehat{U}$ is open and nonempty. By strong approximation, there exists $\beta \in \widehat{U} \cap B^1$. Thus $\operatorname{trd}(\beta\alpha) \in \bigcap_\mathfrak{p} R_\mathfrak{p} = R$ and $\operatorname{nrd}(\beta\alpha) = \operatorname{nrd}(\alpha) = n$. Therefore $\beta\alpha$ is as desired. □

Let $R^\times_{>_\Omega 0} := R^\times \cap F^\times_{>_\Omega 0}$ be the subgroup of units that are positive at the places $v \in \Omega$ (the set of real, ramified places in B).

Corollary 28.6.4. *Suppose B is S-indefinite and that $\mathrm{Cl}_\Omega R$ is trivial. Let $O \subseteq B$ be an Eichler R-order. Then*

$$\operatorname{nrd}(O^\times) = R^\times_{>_\Omega 0}.$$

Proof. Let $u \in R^\times_{>_\Omega 0}$. We repeat the argument of Theorem 28.6.1, with a slight modification. Let $\mathfrak{M}$ be the level of O.

Let $\mathfrak{p} \mid \mathfrak{M}$ be a prime that divides the level $\mathfrak{M}$. We choose an isomorphism $B_\mathfrak{p} \simeq \mathrm{M}_2(F_\mathfrak{p})$ such that $\alpha_\mathfrak{p}$ is in rational canonical form, and let $O'_\mathfrak{p}$ be the standard Eichler order in $\mathrm{M}_2(F_\mathfrak{p})$ of the same level as O. Define

$$U_\mathfrak{p} = \{\beta_\mathfrak{p} \in B_\mathfrak{p}^1 : \beta_\mathfrak{p}\alpha \in O'_\mathfrak{p}\}.$$

This is again an open condition because multiplication is continuous, and the calculation

$$\begin{pmatrix} tu^{-1} & 1 \\ -1 & 0 \end{pmatrix}\begin{pmatrix} 0 & -u \\ 1 & t \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & u \end{pmatrix}$$

shows also that $U_\mathfrak{p} \neq \emptyset$.

For all other primes $\mathfrak{p} \nmid \mathfrak{M}$, we define $U_\mathfrak{p}$ as in (28.6.2). As in the proof of Theorem 28.6.1, we find $\beta \in B^1$ such that $\operatorname{nrd}(\beta\alpha) = u$ and $\gamma' = \beta\alpha$ has $\gamma'_\mathfrak{p} \in O'_\mathfrak{p}$ for all $\mathfrak{p} \mid \mathfrak{M}$. Since γ' is integral, γ' belongs to an R-order O' that is equal to $O'_\mathfrak{p}$ at all $\mathfrak{p} \mid \mathfrak{M}$ and is maximal at all $\mathfrak{p} \nmid \mathfrak{M}$. The order O' is therefore an Eichler order of level $\mathfrak{M}$.

Finally, since $\mathrm{Cl}_{G(O)} R = \mathrm{Cl}_\Omega R$ is trivial, the type set $\operatorname{Typ}(O)$ is also trivial (Corollary 28.5.6), so we conclude $O = \nu O' \nu^{-1}$ for some $\nu \in B^\times$, and $\gamma = \nu\gamma'\nu^{-1} \in O$ has $\operatorname{nrd}(\gamma) = u$ as desired. □

Example 28.6.5. Suppose O is an Eichler order in an indefinite quaternion algebra over $F = \mathbb{Q}$. Then by Corollary 28.6.4, $\operatorname{nrd}(O^\times) = \{\pm 1\}$, in particular, there exists $\gamma \in O^\times$ such that $\operatorname{nrd}(\gamma) = -1$.

Remark 28.6.6. We will prove a stronger version of Corollary 28.6.4 after we have developed the theory of selectivity: see Corollary 31.1.11.

To conclude this section, we consider a variant of principalization of right fractional ideals: we ask further that the generator has totally positive reduced norm.

28.6.7. Suppose F is a number field. Let

$$B_{>0}^{\times} = \{\alpha \in B^{\times} : v(\operatorname{nrd}(\alpha)) > 0 \text{ for all real places } v\}.$$

The reduced norm gives a map $B^{\times}/B_{>0}^{\times} \to F_{>_{\Omega}0}^{\times}/F_{>0}^{\times}$ and the quotient is a finite abelian 2-group, so in particular $B_{>0}^{\times} \le B^{\times}$ has finite index.

Let $I, J \subseteq B$ be R-lattices. We say I, J are **in the same narrow right class** if there exists $\alpha \in B_{>0}^{\times}$ such that $\alpha I = J$; accordingly, we let $\operatorname{Cls}_{\mathsf{R}}^{+} O$ be the **narrow (right) class set** of O. As in Lemma 27.6.8, there is a bijection

$$\operatorname{Cls}_{\mathsf{R}}^{+} O \leftrightarrow B_{>0}^{\times}\backslash \widehat{B}^{\times}/\widehat{O}^{\times}$$

choosing a local generator. The projection map $\operatorname{Cls}_{\mathsf{R}}^{+} O \to \operatorname{Cls}_{\mathsf{R}} O$ has finite fibers as $B_{>0}^{\times} \le B^{\times}$ has finite index, so since $\operatorname{Cls}_{\mathsf{R}} O$ is a finite set, so too is $\operatorname{Cls}_{\mathsf{R}}^{+} O$.

Corollary 28.6.8. *Let F be a number field and suppose B is S-definite. Then the map*

$$\operatorname{nrd}\colon \operatorname{Cls}_{\mathsf{R}}^{+} O \leftrightarrow B_{>0}^{\times}\backslash \widehat{B}^{\times}/\widehat{O}^{\times} \to F_{>0}^{\times}\backslash \widehat{F}^{\times}/\operatorname{nrd}(\widehat{O}^{\times}) =: \operatorname{Cl}_{G(O)}^{+} R \qquad (28.6.9)$$

induced by the reduced norm is a bijection.

Proof. Repeating the argument in the proof of Lemma 28.4.6, we see that the map (28.6.9) is injective for all orders $O' \in \operatorname{Gen} O$ if and only if $\widehat{B}^{1} \subseteq B_{>0}^{\times}\widehat{O}'^{\times}$ for all $O' \in \operatorname{Gen} O$. And the latter holds by strong approximation (Corollary 28.5.12):

$$\widehat{B}^{1} \subseteq B^{1}\widehat{O}'^{1} \subseteq B_{>0}^{\times}\widehat{O}'^{\times}$$

for all R-orders O'. ∎

Proposition 28.6.10. *Let F be a number field. Suppose that B is S-indefinite and that $\operatorname{Cl}_{G(O)}^{+} R = \operatorname{Cl}_{G(O)} R$. For each real place v not in Ω, let $\epsilon_v \in \{\pm 1\}$. Then there exists $\gamma \in O^{\times}$ such that $\operatorname{sgn}(v(\operatorname{nrd}(\gamma))) = \epsilon_v$ for all v real not in Ω.*

In particular, if $F = \mathbb{Q}$ and O is a locally norm-maximal $\mathbb{Z}$-order in an indefinite quaternion algebra B, then there exists $\gamma \in O^{\times}$ with $\operatorname{nrd}(\gamma) = -1$.

Proof. Let $a \in F_{>_{\Omega}0}^{\times}$ be such that $v(a) = \epsilon_v$ for all v real not in Ω and the class of aR is trivial in $\operatorname{Cl}_{G(O)} R$: these constraints together impose congruence conditions on elements in a real cone. By the Hasse–Schilling norm theorem, there exists $\alpha \in B^{\times}$ such that $\operatorname{nrd}(\alpha) = a$. Thus the class of $\operatorname{nrd}(\alpha O)$ in $\operatorname{Cl}_{G(O)}^{+} R = \operatorname{Cl}_{G(O)} R$ is trivial.

But then by (28.6.9) (a consequence of strong approximation), there exists $\beta \in B_{>0}^{\times}$ such that $\beta O = \alpha O$, and therefore $\beta = \alpha\gamma$ with $\gamma \in O^{\times}$. Since β is totally positive, for all real places $v \notin \Omega$ we have $\operatorname{sgn}(v(\operatorname{nrd}(\gamma))) = \operatorname{sgn}(v(\operatorname{nrd}(\alpha))) = \epsilon_v$, completing the proof.

For the second statement, we only need to note that $\operatorname{Cl}^{+} \mathbb{Z} = \operatorname{Cl} \mathbb{Z} = \{1\}$ and recall that $\operatorname{nrd}(O^{\times}) \le \{\pm 1\}$. ∎

Remark 28.6.11. More generally, let B be a central simple algebra over the global field F. We say B **satisfies the S-Eichler condition** if there exists a place $v \in S$ such that B_v is not a division algebra. (In this text, for quaternion algebras we prefer to use the term S-*indefinite* because it readily conveys the notion, but both terms are common.) If F is a number field and S is the set of archimedean places of F, then B satisfies the S-Eichler condition if and only if B is *not* a totally definite quaternion algebra (Exercise 28.5). So the condition is a mild condition, and the quaternion algebra case requires special effort.

When B satisfies the S-Eichler condition, then $B^1 \hookrightarrow \widehat{B}^1$ is dense, and for $O \subseteq B$ a maximal R-order, a locally principal right fractional O-ideal $I \subseteq B$ is principal if and only if its reduced norm $\operatorname{nrd}(I) \subseteq R$ is trivial in $\operatorname{Cl}_\Omega R$, where Ω is the set of real places $v \in \operatorname{Pl} F$ such that $B_v \simeq \operatorname{M}_n(\mathbb{H})$, generalizing the quaternion case.

Eichler proved Theorem 28.5.5 and the more general statement of the previous paragraph [Eic37, Satz 2], also providing several reformulations and applications [Eic38a, Eic38c]. For a full exposition, see Reiner [Rei2003, §34].

28.7 First proof

Now we proceed with the proof of strong approximation in Theorem 28.5.3; we follow roughly the same lines as in the proof of Eichler's theorem on norms, but here instead we will be concerned with traces. In other words, we replace strong approximation of elements by strong approximation of traces, and then we just have to chase conjugacy classes.

We start with a statement of weak and strong approximation for the global field F.

Lemma 28.7.1 (Weak approximation for F). *Let $S \subseteq \operatorname{Pl} F$ be a finite nonempty set of places. Then the images of the maps*

$$F \hookrightarrow F_S := \prod_{v \in S} F_v \quad \textit{and} \quad F^\times \hookrightarrow F_S^\times := \prod_{v \in S} F_v^\times$$

are dense.

Proof. See e.g. Neukirch [Neu99, Theorem II.3.4] or O'Meara [O'Me73, §11E]. □

Lemma 28.7.2 (Strong approximation for F). *Let $S \subseteq \operatorname{Pl} F$ be a finite nonempty set of places. Then the image of $F \hookrightarrow \underline{F}_{\not S} := \prod_{v \notin S} F_v$ is dense.*

Proof. See e.g. Neukirch [Neu99, Exercise III.1.1] or O'Meara [O'Me73, §33G]. □

We recall that B is a quaternion algebra over F and we write $B_S := \prod_{v \in S} B_v$.

Proposition 28.7.3 (Weak approximation for B). *Let $S \subseteq \operatorname{Pl} F$ be a finite nonempty set of places. Then the images*

$$B \hookrightarrow B_S \quad \textit{and} \quad B^\times \hookrightarrow B_S^\times \quad \textit{and} \quad B^1 \hookrightarrow B_S^1$$

are dense.

Proof. By weak approximation for F (Lemma 28.7.1), we have F dense in $\prod_{v \in S} F_v$. Choosing an F-basis for B, we have $B \simeq F^4$ as topological F-vector spaces, and so by approximating in each coordinate, we conclude that B is dense in $\prod_{v \in S} B_v$. The multiplicative case follows *a fortiori*, restricting open neighborhoods.

Finally we treat B^1. By Exercise 7.31, we know that $B^1 = [B^\times, B^\times]$ is the commutator. Let $(\gamma_v)_v \in \prod_{v \in S} B_v^1$. Then for each $v \in S$, we can write $\gamma_v = \alpha_v \beta_v \alpha_v^{-1} \beta_v^{-1}$ with $\alpha_v, \beta_v \in B^\times$. Then by weak approximation for $B^\times$, we can find a sequence $\alpha_n \in B^\times$ such that $\alpha_n \to (\alpha_v)_v \in B_S^\times$, and similarly with $\beta_n \to (\beta_v)_v$. Then since multiplication is continuous, we conclude that $\gamma_n = \alpha_n \beta_n \alpha_n^{-1} \beta_n^{-1} \to (\gamma_v)_v \in B^1$. □

Next, we need to approximate polynomials: this kind of lemma was first performed in section 14.7 to prove the Hasse–Schilling theorem of norms, and here we need another variant.

Lemma 28.7.4. *Let $S \subseteq \operatorname{Pl} F$ be a finite nonempty set of places and $\Sigma \subseteq \operatorname{Pl} F$ a finite set of noncomplex places disjoint from S. Let $\widehat{t} \in \widehat{B}$ be such that $x^2 - t_v x + 1 \in F_v[x]$ is irreducible for $v \in \Sigma$.*

Let $\epsilon > 0$. Then there exists $t \in F$ such that:

- *$f(x) = x^2 - tx + 1$ is irreducible and separable over F;*
- *$|t - t_v|_v < \epsilon$ for all $v \in \Sigma$;*
- *$f(x)$ is irreducible over F_v for all $v \in \Sigma$; and*
- *$|t - t_v|_v \leq 1$ for all $v \notin S \cup \Sigma$.*

Proof. We argue as in Lemma 14.7.6 (and Corollary 14.7.8), but instead of weak approximation we now use strong approximation (Lemma 28.7.2). Our job is a bit easier because we are only asking for irreducibility, not separability.

Since S is nonempty, by strong approximation for F we can find t arbitrarily close to $\widehat{t}$, thus ensuring that the desired inequalities hold and that $f(x)$ is irreducible over F_v for $v \in \Sigma$; to ensure that $f(x)$ is separable, we need only avoid the locus $t^2 = 4$, and similarly we may ensure $f(x)$ is irreducible. □

We now embark on the proof of strong approximation.

Proof of Main Theorem 28.5.3. We follow Vignéras [Vig80a, Théorème III.4.3]; see also Miyake [Miy2006, Theorems 5.2.9–5.2.10] for the case $F = \mathbb{Q}$. We show that the closure of B^1 is equal to $\widehat{B}^1$. Let $\widehat{\gamma} = (\gamma_v)_v \in \widehat{B}^1$; we will find a sequence of elements of B^1 converging to $\widehat{\gamma}$.

Step 1: Setup. We claim it is enough to consider the case where $\gamma_v = 1$ for all but finitely many v, by a Cantor-style diagonalization argument. Indeed, for a finite set $T \subseteq \operatorname{Pl} F$ disjoint from S, we let $\widehat{\gamma}_{[T]}$ be the idele obtained from $\widehat{\gamma}$ replacing $\gamma_v = 1$ for $v \notin T$. Then for a sequence of subsets T eventually containing each place v, we have $\widehat{\gamma}_{[T]} \to \widehat{\gamma}$. Thus, if we can find sequences in B^1 converging to each $\widehat{\gamma}_{[T]}$ we can diagonalize to find a sequence converging to $\widehat{\gamma}$, since $\operatorname{Pl} F$ is countable.

So we may suppose without loss of generality that $\gamma_v = 1$ for all but finitely many v. To find a sequence in B^1 converging to $\widehat{\gamma}$, our strategy in the proof is as follows: in shrinking open neighborhoods of $\widehat{\gamma}$ we first find an element in B^1 whose reduced

characteristic polynomial is close to an element in the open neighborhood, and then we conjugate by $B^\times$ to get the limits themselves to match.

To this end, let $O \subset B$ be a reference R-order, let $T \subseteq \mathrm{Pl}\, F$ be a finite set of places disjoint from S containing the primes v where $\gamma_v \neq 1$ and the ramified places of B not in S. We consider open neighborhoods of the form

$$U = \prod_{v \in T} \gamma_v U_v \times \prod_{v \notin S \cup T} O_v^1$$

where $U_v \subseteq B_v^1$ is an open neighborhood of 1.

Step 2: Polynomial approximation. Now comes the polynomial approximation step: we will show that there exists $\widehat{\alpha} \in \widehat{B}^\times$ such that $B^1 \cap \widehat{\alpha}^{-1} U \widehat{\alpha} \neq \emptyset$. (This is about as good as could be expected at this stage: if we argue by approximating polynomials, we should only be able to expect to get something up to conjugation.)

We define the idele $\widehat{t}$ in each component, as follows.

- If $v \notin S \cup T$, we take $t_v = \mathrm{trd}(\gamma_v) = 2$.
- If $v \in T$ is unramified in B, we take $t_v = \mathrm{trd}(\gamma_v)$.
- If $v \in T$ is ramified in B, we choose $\mu_v \in \gamma_v U_v$ such that $\mu_v \notin F_v$ is separable over F_v, and take $t_v = \mathrm{trd}(\mu_v)$. Since B_v is a division algebra, we have $F_v[\mu_v]$ is a field, and so its reduced characteristic polynomial is irreducible.
- If $v \in S$ is ramified in B, we choose $t_v \in F_v$ such that $x^2 - t_v x + 1$ is irreducible; such an element t_v exists by Lemma 14.7.5).

The only places that remain are those $v \in S$ such that v is unramified in B. By hypothesis that B is S-indefinite, we know that there is at least one such split place $v_{\mathrm{spl}} \in S$ remaining (and in particular, $v_{\mathrm{spl}} \notin \mathrm{Ram}\, B \cup T$).

We now apply our polynomial approximation Lemma 28.7.4, we conclude that there exists $t \in F$ such that $f(x) = x^2 - tx + 1$ is separable and irreducible over F, and irreducible over F_v for all $v \in \mathrm{Ram}\, B$, and such that t is arbitrarily close to $\widehat{t}$. Let $K = F[x]/(f(x))$. Then either $\mathrm{Ram}\, B = \emptyset$ and $K \hookrightarrow B$ automatically, or $\mathrm{Ram}\, B \neq \emptyset$ so $f(x)$ is irreducible and then $K \hookrightarrow B$ by the local-global principle for embeddings (Proposition 14.6.7). Let $\beta \in B^1$ have minimal polynomial $f(x)$. Since $\widehat{t} \in \mathrm{trd}(U)$ and the reduced trace is an open (linear) map, with a closer approximation we may suppose $\mathrm{trd}(\beta) \in \mathrm{trd}(U)$, and therefore there exists $\widehat{\gamma}' \in U$ such that $\mathrm{trd}(\beta) = \mathrm{trd}(\widehat{\gamma}')$ so that $\beta, \widehat{\gamma}'$ have the same irreducible minimal polynomial. By the Skolem–Noether theorem (Corollary 7.7.3), there exists $\widehat{\alpha} \in \widehat{B}^\times$ such that

$$\beta = \widehat{\alpha}^{-1} \widehat{\gamma}' \widehat{\alpha}. \tag{28.7.5}$$

Step 3: Finding a sequence. We then repeat the above argument with a sequence of open neighborhoods $U_n \ni \widehat{\alpha}$ such that $\bigcap_n U_n = \{\widehat{\gamma}\}$; we obtain a sequence

$$B^1 \ni \beta_n = \widehat{\alpha}_n^{-1} \widehat{\gamma}_n' \widehat{\alpha}_n \in \widehat{\alpha}_n^{-1} U_n \widehat{\alpha}_n. \tag{28.7.6}$$

Since $\widehat{\gamma}_n' \in U_n$, we have $\widehat{\gamma}_n' \to \widehat{\gamma}$, and in particular for $v \in \mathrm{Pl}\, F \smallsetminus S \cup T$, we have $\gamma_{n,v}' \to \gamma_v = 1$.

Step 4: Harmonizing the sequence. By 'harmonizing' the conjugating elements $\widehat{\alpha}_n$, we will realize a sequence in B^1 tending to $\widehat{\gamma}$ as desired, in two (subset)steps. First, by Main Theorem 27.6.14, $\widehat{B}^\times/B^\times$ is compact. So restricting to a subsequence, we have $\widehat{\alpha}_n = \widehat{\delta}_n\mu_n$ with $\mu_n \in B^\times$ and $\widehat{\delta}_n \to \widehat{\delta} = (\delta_v)_v \in \widehat{B}^\times$. Second, by weak approximation for B (Proposition 28.7.3), $B^\times$ is dense in $\prod_{v\in T} B_v^\times$, so there is a sequence ν_n from $B^\times$ such that $\nu_n \to (\delta_v)_{v\in T}$.

Step 5: Conclusion. To conclude, we consider the sequence

$$(\nu_n\mu_n)\beta_n(\nu_n\mu_n)^{-1} = (\nu_n\widehat{\delta}_n^{-1})\widehat{\gamma}'_n(\widehat{\delta}_n\nu_n^{-1}). \tag{28.7.7}$$

We claim that this sequence tends to $\widehat{\gamma}$. For $v \in T$, we have $\nu_{n,v}\widehat{\delta}_{n,v}^{-1} \to \delta_v\delta_v^{-1} = 1$ so

$$(\nu_{n,v}\delta_{n,v}^{-1})\gamma'_{n,v}(\delta_{n,v}\nu_{n,v}^{-1}) \to \gamma_v. \tag{28.7.8}$$

On the other hand, for $v \in \mathrm{Pl}\,F\smallsetminus(S\cup T)$, we have $\gamma'_{n,v} \to 1$, so

$$(\nu_{n,v}\delta_{n,v}^{-1})\gamma'_{n,v}(\delta_{n,v}\nu_{n,v}^{-1}) \to 1 = \gamma_v. \tag{28.7.9}$$

Putting together these cases, we conclude that $(\nu_n\mu_n)\beta_n(\nu_n\mu_n)^{-1} \to \widehat{\gamma}$, and therefore $\widehat{\gamma}$ is in the closure of B^1. □

Remark 28.7.10. Strong approximation has a more general formulation, as follows. Let G be a semisimple, simply-connected algebraic group over the global field F. Let S be an eligible set containing a place v such that $G(F_v)$ is not compact. Then $G(F)$ is dense in $G(\underline{F}_{\$})$, and we say G **satisfies strong approximation** (relative to S). Over number fields, strong approximation was established by Kneser [Kne66a, Kne66b] and Platonov [Pla69, Pla69-70], and over function fields by Margulis and Prasad [Pra77]; see also Platonov–Rapinchuk [PR94, Theorem 7.12]. For a survey with discussion and bibliography, see Rapinchuk [Rap2014] and the description of Kneser's work by Scharlau [Scha2009, §2.1].

28.8 Second proof

Because of its importance, we now give a second proof of strong approximation; this has the same essential elements, but uses some facts from group theory to simplify away the final steps of the previous proof. We follow Swan [Swa80, §14], who references Kneser [Kne66a, Kne66b], Platonov [Pla69-70], and Prasad [Pra77]; see also the exposition by Kleinert [Klt2000, §4.2].

Let $Z := \mathrm{cl}(B^1) \le \widehat{B}^1$ be the closure of B^1 in $\widehat{B}^1$. We embed $B_v^1 \hookrightarrow \widehat{B}^1$ by $(\dots, 1, \alpha_v, 1, \dots)$ in the vth component, for $v \notin S$.

Lemma 28.8.1. *If $B_v^1 \subseteq Z$ for almost all $v \notin S$, then $Z = \widehat{B}^1$.*

Proof. Suppose that $B_v^1 \subseteq Z$ for all $v \notin T$ where T is a finite set. Let $\widehat{\gamma} \in \widehat{B}^1$. If $\gamma_v = 1$ for all $v \in T$, then $\widehat{\gamma}$ is a limit of elements in Z (approximating at a finite level), so $\widehat{\gamma} \in Z$. Otherwise, by weak approximation for B (Proposition 28.7.3), there exists $\gamma \in B^1$ such that γ is near γ_v for all $v \in T$. Let $\widehat{\beta}$ have $\beta_v = 1$ for $v \in T$ and $\beta_v = \gamma^{-1}\gamma_v$ for $v \notin T$; then $\widehat{\beta} \in Z$, and $\widehat{\gamma}$ is the limit of the $\gamma\widehat{\beta}$. □

Now we consider

$$Z_1 := \{\widehat{\gamma} \in Z : \gamma_v = 1 \text{ for all but finitely many } v\}. \qquad (28.8.2)$$

Lemma 28.8.3. *$Z_1 \trianglelefteq \widehat{B}^1$ is a normal subgroup.*

Proof. Let $\widehat{\gamma} \in \widehat{B}^1$ and let $\widehat{\alpha} \in Z_1$ with $\alpha_v = 1$ for $v \notin T$ with T a finite set. By weak approximation for B (Proposition 28.7.3), there exists $\gamma \in B^1$ with γ close to γ_v for all $v \in T$. Therefore $\gamma^{-1}\widehat{\alpha}\gamma$ is near $\widehat{\gamma}^{-1}\widehat{\alpha}\widehat{\gamma}$ for $v \in T$ and $\gamma^{-1}\alpha_v\gamma = \gamma_v^{-1}\alpha_v\gamma_v = 1$ for $v \notin T$, so is the limit of such in Z. Therefore $\widehat{\gamma}^{-1}\widehat{\alpha}\widehat{\gamma} \in Z_1$, thus $\widehat{\gamma}^{-1}Z_1\widehat{\gamma} \subseteq Z_1$ and $Z_1 \trianglelefteq \widehat{B}^1$. □

Lemma 28.8.4. *Let F be an infinite field. Then $\mathrm{PSL}_2(F)$ is a simple group.*

Proof. See e.g. Grove [Grov2002, Theorem 1.13]. Briefly, the result can be proven using Iwasawa's criterion, since $\mathrm{SL}_2(F)$ acts doubly transitively on the linear subspaces of F^2: the kernel of the action is the center $\{\pm 1\}$, and the stabilizer subgroup of a standard basis element is the subgroup of upper triangular matrices, whose conjugates generate $\mathrm{SL}_2(F)$. □

Proof of Main Theorem 28.5.3 (Strong approximation). We show that $Z = \widehat{B}^1$. By Lemma 28.8.1, it is enough to show that $B_v^1 \subseteq Z$ for almost all v. By Lemma 28.8.3, each $Z_1 \cap B_v^1 \trianglelefteq B_v^1$ is a normal subgroup; by Lemma 28.8.4, either this normal subgroup is either scalar, or we have $Z_1 \cap B_v^1 = B_v^1 \le Z_1 \le Z$ and we are done. So it suffices to show that for almost all v, we have $Z_1 \cap B_v^1$ nonscalar, which is to say, $Z_1 \cap B_v^1 \neq \{\pm 1\}$.

So let $w \in \mathrm{Pl}\, F$ be unramified. We perform polynomial approximation, as in Step 2 of the first proof. For $v \in \mathrm{Ram}\, B$, let $x^2 - t_v x + 1$ be a separable irreducible polynomial (which exists by Lemma 14.7.5) with $t_v \in R_v$, and do the same for w; let $\widehat{t}$ be the corresponding idele, with $t_v = 1$ for the remaining places v. By Lemma 28.7.4, there exists $t \in F$ such that $f(x) = x^2 - tx + 1$ is:

- irreducible and separable over F,
- irreducible over F_v for all $v \in \mathrm{Ram}\, B$, and
- such that t arbitrarily well-approximates $\widehat{t} \in \widehat{R}$, so we may suppose $t \in R$.

By the local-global principle for embeddings (Proposition 14.6.7), there exists $\beta \in B^1$ with $f(\beta) = 0$; but moreover, β is integral and so belongs to a maximal order. In this way, we manufacture a sequence β_n with $\mathrm{trd}(\beta_n) \to \widehat{t}$. Repeating this with $t_v \to 1$ for $v \in \mathrm{Ram}\, B$ and diagonalizing, we may suppose $t_v = 1$ for all $v \neq w$.

Let O be a maximal R-order. The type set $\mathrm{Typ}\, O$ is finite by Corollary 27.6.25; choose representatives O_i for $\mathrm{Typ}\, O$. After conjugating the elements β_n, we may suppose without loss of generality that each β_n belongs to one of the orders O_i. By the pigeonhole principle, there is an order containing infinitely many, so restricting to a subsequence we may suppose $\beta_n \in O^1$ for all n.

But now the kicker: $\widehat{O}^1$ is compact, so we may restrict to a convergent subsequence $\beta_n \to \widehat{\beta} \in \widehat{O}^1$. By construction, each β_n has separable reduced characteristic polynomial, and $\mathrm{trd}(\beta_{n,v}) \to 1$ for all $v \neq w$, so $\beta_{n,v} \to \beta_v = 1$ for all $v \neq w$. But $\mathrm{trd}(\beta_{n,w}) \to t_w$, and $x^2 - t_w x + 1$ is irreducible, so $\beta_w \notin F_w^\times$, as desired. □

28.9 ∗ Normalizer groups

In this section, we apply strong approximation to the normalizer of an order, and we compare the normalizers for locally isomorphic orders. We recall the notation from section 18.5. Let $\mathrm{Idl}(O)$ be the group of invertible two-sided fractional O-ideals, let $\mathrm{PIdl}(O) \leq \mathrm{Idl}(O)$ the subgroup of principal fractional O-ideals, and let $\mathrm{PIdl}(R) \leq \mathrm{PIdl}(O)$ be the image of the group of principal fractional R-ideals.

We suppose throughout this section that $\mathrm{PIdl}(O) \trianglelefteq \mathrm{Idl}(O)$ is a normal subgroup. This is true whenever $\mathrm{Idl}\, O$ is abelian, which holds when O is Eichler order (using Lemma 23.3.13 for the primes where $\mathfrak{p}$ is maximal and Proposition 23.4.14 the remaining primes).

28.9.1. Recall there is a natural exact sequence

$$1 \to N_{B^\times}(O)/(F^\times O^\times) \to \mathrm{Pic}_R(O) \to \mathrm{Idl}(O)/\,\mathrm{PIdl}(O) \to 1 \qquad (18.5.5)$$

obtained by considering the class of a bimodule as a two-sided ideal modulo principal ideals. The idelic dictionary (27.6.26) gives another proof of exactness: there is a canonical bijection

$$\mathrm{Idl}(O) \xrightarrow{\sim} N_{\widehat{B}^\times}(\widehat{O})/\widehat{O}^\times,$$

to obtain $\mathrm{Idl}(O)/\,\mathrm{PIdl}(O)$ we take the quotient by $N(O)$ and to obtain $\mathrm{Idl}(O)/\,\mathrm{PIdl}(R)$ we take the quotient by $F^\times$; therefore the exact sequence (18.5.5) can be rewritten

$$1 \to F^\times\backslash N_{B^\times}(O)/O^\times \to F^\times\backslash N_{\widehat{B}^\times}(\widehat{O})/\widehat{O}^\times \to N_{B^\times}(O)\backslash N_{\widehat{B}^\times}(\widehat{O})/\widehat{O}^\times \to 1 \qquad (28.9.2)$$

and its exactness is now visible.

28.9.3. We have a map of pointed sets

$$\begin{aligned} \mathrm{Idl}(O) &\to \mathrm{Cls}\, O \\ I &\mapsto [I] \end{aligned}$$

and $\mathrm{PIdl}(O)$ is the kernel of this map, the preimage of the trivial class in $\mathrm{Cls}\, O$. The composition of this map with the reduced norm gives a group homomorphism:

$$\begin{aligned} c\colon\ \mathrm{Idl}(O) &\to \mathrm{Cl}_{G(O)}(R) = F^\times_{>_\Omega 0}\backslash \widehat{F}^\times/\,\mathrm{nrd}(\widehat{O}) \\ I &\mapsto [\mathrm{nrd}(I)] \end{aligned}$$

Lemma 28.9.4. *Suppose that B is S-indefinite. Then* $\mathrm{PIdl}(O) = \ker c$*, i.e.,*

$$\mathrm{PIdl}(O) = \{I \in \mathrm{Idl}(O) : [\mathrm{nrd}(I)] \textit{ is trivial in } \mathrm{Cl}_{G(O)}(R)\}.$$

Proof. By strong approximation (Theorem 28.5.5), the reduced norm gives a bijection $\mathrm{nrd}\colon\ \mathrm{Cls}\, O \to \mathrm{Cl}_{G(O)}(R)$; thus $I \in \mathrm{Idl}(O)$ is principal, belonging to $\mathrm{PIdl}(O)$, if and only if $[\mathrm{nrd}(I)]$ is trivial in $\mathrm{Cl}_{G(O)}(R)$. □

Now let O, O' be locally isomorphic orders (in the same genus) with connecting O, O'-ideal J.

28.9.5. By 18.4.7, there is an isomorphism of groups

$$\begin{aligned} \mathrm{Idl}(O) &\xrightarrow{\sim} \mathrm{Idl}(O') \\ I &\mapsto J^{-1}IJ. \end{aligned} \tag{28.9.6}$$

which induces an isomorphism $\mathrm{Pic}_R(O) \simeq \mathrm{Pic}_R(O')$.

We now come to the first major result of this section.

Proposition 28.9.7. *Suppose that B is S-indefinite. Then the map* (28.9.6) *induces a commutative diagram*

$$\begin{array}{ccccccccc} 1 & \longrightarrow & N_{B^\times}(O)/(F^\times O^\times) & \longrightarrow & \mathrm{Pic}_R(O) & \longrightarrow & \mathrm{Idl}(O)/\,\mathrm{PIdl}(O) & \longrightarrow & 1 \\ & & \downarrow\wr & & \downarrow\wr & & \downarrow\wr & & \\ 1 & \longrightarrow & N_{B^\times}(O')/(F^\times O'^\times) & \longrightarrow & \mathrm{Pic}_R(O') & \longrightarrow & \mathrm{Idl}(O')/\,\mathrm{PIdl}(O') & \longrightarrow & 1 \end{array}$$

with vertical maps isomorphisms.

Proof. We verify that $J^{-1}\,\mathrm{PIdl}(O)J = \mathrm{PIdl}(O')$, from which both statements follow; and this verification comes from Lemma 28.9.4, as

$$[\mathrm{nrd}(J^{-1}IJ)] = [\mathrm{nrd}(I)] \in \mathrm{Cl}_{G(O)}\, R = \mathrm{Cl}_{G(O')}\, R$$

(recall 27.7.4), so $I \in \mathrm{PIdl}(O)$ if and only if $J^{-1}IJ \in \mathrm{PIdl}(O')$. ☐

Proposition 28.9.7 says that when B is S-indefinite, then the structure of the normalizer group, the Picard group, and group of ideals modulo principal ideals are all isomorphic for all orders in a genus. The same is *not* true when B is S-definite; we always have an isomorphism in the middle, but for locally isomorphic orders, the Picard group may be distributed differently between the normalizer and the ideal group.

By chasing a few diagrams, we can be more specific about the structure of $\mathrm{Idl}(O)$ by seeking out primitive ideals.

Definition 28.9.8. The **Atkin–Lehner group** of O is

$$\mathrm{AL}(O) := \left\{J \in \mathrm{Idl}(O) : [\mathrm{nrd}(J)] \in (\mathrm{Cl}_{G(O)}\, R)^2\right\}/\,\mathrm{Idl}(R). \tag{28.9.9}$$

The definition (28.9.9) makes sense because $\mathrm{Idl}(R)$ is indeed a subgroup: since if $\mathfrak{a} \in \mathrm{Idl}(R)$ then $[\mathrm{nrd}(\mathfrak{a}O)] = [\mathfrak{a}]^2 \in (\mathrm{Cl}_{G(O)}\, R)^2$.

Example 28.9.10. Suppose that O is an Eichler order with $\mathrm{discrd}\, O = \mathfrak{N}$. Then (23.4.21) gives an isomorphism

$$\mathrm{Idl}(O)/\,\mathrm{Idl}(R) \simeq \prod_{\mathfrak{p} \mid \mathfrak{N}} \mathbb{Z}/2\mathbb{Z}.$$

The group $\mathrm{AL}(O)$ is therefore an abelian 2-group, isomorphic to $\prod_{\mathfrak{p}\mid\mathfrak{N}} \mathbb{Z}/2\mathbb{Z}$ when $(\mathrm{Cl}_{G(O)}\, R)^2$ is trivial. For example, for $O = \mathrm{M}_2(R) \subset B = \mathrm{M}_2(F)$, we have $\mathrm{AL}(O)$ the trivial group.

28.9.11. There is a group homomorphism

$$\begin{aligned} \operatorname{Idl}(R) &\to \operatorname{Idl}(O) \\ \mathfrak{a} &\mapsto \mathfrak{a}O; \end{aligned} \tag{28.9.12}$$

this map is injective, since $\mathfrak{a}O = O$ implies $\mathfrak{a} = R$. We obtain an exact sequence

$$1 \to \operatorname{Cl} R \to \operatorname{Pic}_R(O) \to \operatorname{Idl}(O)/\operatorname{Idl}(R) \to 1 \tag{28.9.13}$$

(compare to (18.5.5)). From Lemma 28.9.4 and the fact that $\operatorname{PIdl}(R) \subseteq \ker(c)$, we obtain an exact sequence

$$1 \to N_{B^\times}(O)/(F^\times O^\times) \to \operatorname{Pic}_R(O) \xrightarrow{c} \operatorname{Cl}_{G(O)}(R). \tag{28.9.14}$$

From (28.9.14), we see that $c(\operatorname{Idl}(R)/\operatorname{PIdl}(R)) = (\operatorname{Cl}_{G(O)} R)^2$; further, we have $c(\operatorname{Cl} R) = (\operatorname{Cl}_{G(O)} R)^2$ with

$$\ker c|_{\operatorname{Cl} R} = (\operatorname{Cl} R)[2]_{\uparrow O} := \operatorname{img}\bigl(\operatorname{Cl}_{G(O)}(R)[2] \to \operatorname{Cl}(R)\bigr) \leq (\operatorname{Cl} R)[2]. \tag{28.9.15}$$

We write $(\operatorname{Cl} R)[2]_{\uparrow O}$ because this is the subgroup of ideal classes that lift to the group $\operatorname{Cl}_{G(O)}(R)[2]$ (having order dividing 2). Therefore the following diagram commutes, with exact rows and columns:

$$\begin{array}{ccccccccc}
 & & 1 & & 1 & & 1 & & \\
 & & \downarrow & & \downarrow & & \downarrow & & \\
1 & \to & (\operatorname{Cl} R)[2]_{\uparrow O} & \to & N_{B^\times}(O)/(F^\times O^\times) & \to & \operatorname{AL}(O) & & \\
 & & \downarrow & & \downarrow & & \downarrow & & \\
1 & \to & \operatorname{Cl} R & \to & \operatorname{Pic}_R(O) & \to & \operatorname{Idl}(O)/\operatorname{Idl}(R) & \to & 1 \\
 & & \downarrow & & \downarrow{\scriptstyle c} & & \downarrow & & \\
1 & \to & (\operatorname{Cl}_{G(O)} R)^2 & \to & \operatorname{Cl}_{G(O)}(R) & \to & \operatorname{Cl}_{G(O)}(R)/(\operatorname{Cl}_{G(O)} R)^2 & \to & 1 \\
 & & \downarrow & & & & & & \\
 & & 1 & & & & & &
\end{array} \tag{28.9.16}$$

The second main result of this section is the following.

Proposition 28.9.17. *Suppose B is S-indefinite. Then there is a (non-canonically) split exact sequence*

$$1 \to (\operatorname{Cl} R)[2]_{\uparrow O} \to N_{B^\times}(O)/(F^\times O^\times) \to \operatorname{AL}(O) \to 1. \tag{28.9.18}$$

Proof. The snake lemma implies that the top row of (28.9.16) is exact; the sequence is split by choosing for each class in $\operatorname{AL}(O)$ a generator of a representative ideal.

Here is second, self-contained proof which captures the above discussion. Say that a two-sided (integral) O-ideal I is R-**primitive** if I is not divisible by any integral ideal of the form $\mathfrak{a}O$ with $\mathfrak{a} \subsetneq R$. (If I is integral but not R-primitive, with $I \subseteq \mathfrak{a}O$ and $\mathfrak{a}$ as

small as possible, then $\mathfrak{a}^{-1}I \subseteq O$ is integral and now R-primitive.) Let $\alpha \in N_{B^\times}(O)$. Then $O\alpha O = I \in \mathrm{Idl}(O)$, so we can factor $I = \mathfrak{c}J$ uniquely with $\mathfrak{c}$ a fractional ideal of R and J an R-primitive ideal. We have

$$aR = \mathrm{nrd}(\alpha)R = \mathrm{nrd}(I) = \mathfrak{c}^2 \,\mathrm{nrd}(J) = \mathfrak{c}^2\mathfrak{a} \tag{28.9.19}$$

and so

$$1 = [(a)] = [\mathfrak{c}]^2[\mathfrak{a}] \in \mathrm{Cl}_{G(O)}(R)$$

and in particular $[\mathrm{nrd}(J)] \in (\mathrm{Cl}_{G(O)} R)^2$. Therefore there is a map

$$N_{B^\times}(O) \to \mathrm{AL}(O). \tag{28.9.20}$$

This map is surjective by strong approximation (see Lemma 28.9.4): if $J \in \mathrm{Idl}(O)$ has $\mathrm{nrd}(J) = \mathfrak{a}$ and $[\mathfrak{a}] = [\mathfrak{c}^{-1}]^2 \in (\mathrm{Cl}_{G(O)} R)^2$, then $[\mathrm{nrd}(\mathfrak{c}J)] = 1 \in \mathrm{Cl}_{G(O)}(R)$ so by Theorem 28.5.5, there exists $\alpha \in B^\times$ such that $O\alpha O = \mathfrak{c}J$ and since $\mathfrak{c}J \in \mathrm{Idl}(O)$ we have $\alpha \in N_{B^\times}(O)$. The map is split by this construction, with a choice of J up to $\mathrm{Idl}\,R$. The kernel of the map in (28.9.20) consists of $\alpha \in N_{B^\times}(O)$ such that $O\alpha O = \mathfrak{c}O$ with $\mathfrak{c} \in \mathrm{Idl}(R)$, and from the preceding paragraph $[\mathfrak{c}^2] = 1 \in \mathrm{Cl}_{G(O)}(R)$ so $[\mathfrak{c}] \in \mathrm{Cl}_{G(O)}(R)[2]$; however, the kernel also contains $F^\times O^\times$, whence the class of $\mathfrak{c}$ is well-defined only in $\mathrm{Cl}(R)$. Therefore the kernel is canonically identified with $(\mathrm{Cl}\,R)[2]_{\uparrow O}$. ◻

Corollary 28.9.21. *Suppose that O is an Eichler order with* $\mathrm{discrd}\,O = \mathfrak{N}$ *and that B is S-indefinite. Then*

$$N_{B^\times}(O)/(F^\times O^\times) \simeq \mathrm{AL}(O) \times (\mathrm{Cl}\,R)[2]_{\uparrow O}$$

is an abelian 2-group with rank at most $\omega(\mathfrak{N}) + h_2(R)$, where $\omega(\mathfrak{N})$ is the number of prime divisors of $\mathfrak{N}$ and $h_2(R) = \dim_{\mathbb{F}_2}(\mathrm{Cl}\,R)[2]$.

Proof. Combine Proposition 28.9.17 and Example 28.9.10. ◻

Corollary 28.9.22. *We have*

$$N_{\mathrm{GL}_2(F)}(\mathrm{M}_2(R))/(F^\times \mathrm{GL}_2(R)) \simeq (\mathrm{Cl}\,R)[2].$$

Proof. Apply Corollary 28.9.21 with $\mathrm{Cl}_{G(O)}(R) = \mathrm{Cl}(R)$ so $(\mathrm{Cl}\,R)[2]_{\uparrow O} = (\mathrm{Cl}\,R)[2]$ and $\mathrm{AL}(O)$ the trivial group by Example 28.9.10. ◻

Remark 28.9.23. Corollary 28.9.21 corrects Vignéras [Vig80a, Exercise III.5.4] to account for possible class group factors.

28.10 ∗ Stable class group

We conclude this epic chapter with a final result on the stable class group, restoring some generality; we announced a special case of this theorem as Theorem 20.7.16.

Theorem 28.10.1 (Fröhlich–Swan). *Let R be a global ring with $F = \mathrm{Frac}\,R$, let B be a central simple F-algebra, and let $O \subset B$ be an R-order. Let $\Omega \subseteq \mathrm{Pl}\,F$ be the set of real*

places v of F such that $B_v \simeq \mathrm{M}_m(\mathbb{H})$ for some $m \geq 1$. Then the reduced norm induces an isomorphism

$$\mathrm{nrd}\colon \operatorname{StCl} O \xrightarrow{\sim} F^\times_{>_\Omega 0}\backslash \widehat{F}^\times / \mathrm{nrd}(\widehat{O}^\times) \tag{28.10.2}$$

of finite abelian groups. In particular, if B is a quaternion algebra and O is locally norm-maximal order, then

$$\operatorname{StCl} O \simeq \mathrm{Cl}_\Omega R$$

where $\Omega \subseteq \operatorname{Ram} B$ is the set of real ramified places in B.

Proof. We give only a sketch of the proof. For further details, see the references given for the proof of Theorem 20.7.16.

We first show that the map (28.10.2) is well-defined. Suppose that $I \oplus O^r \simeq I' \oplus O^r$ with $r \geq 0$. If $r = 0$, we are done; so suppose $r \geq 1$. Extending scalars, we find an isomorphism $\phi\colon B^{r+1} \to B^{r+1}$ of left B-modules, represented by an element $\gamma \in \mathrm{GL}_{r+1}(B)$ acting on the left. In a similar way, associated to $I \oplus O^r$ we obtain a class

$$\mathrm{GL}_{r+1}(B)\widehat{\alpha}\,\mathrm{GL}_{r+1}(\widehat{O}) \in \mathrm{GL}_{r+1}(B)\backslash \mathrm{GL}_{r+1}(\widehat{B}) / \mathrm{GL}_{r+1}(\widehat{O})$$

by choosing in each completion an isomorphism with $O_{\mathfrak{p}}^{r+1}$ represented by a matrix, well-defined up to a change of basis on the right (and on the left by a global isomorphism). Now by strong approximation in the advanced version announced in Remark 28.7.10, the reduced norm induces a bijection

$$\mathrm{nrd}\colon \mathrm{GL}_{r+1}(B)\backslash \mathrm{GL}_{r+1}(\widehat{B}) / \mathrm{GL}_{r+1}(\widehat{O}) \to F^\times_{>_\Omega 0}\backslash \widehat{F}^\times / \mathrm{nrd}(\widehat{O}^\times)$$

after we check that the codomain is indeed the image of the reduced norm. Then

$$\mathrm{nrd}(\widehat{\alpha}') = \mathrm{nrd}(\gamma\widehat{\alpha}) = \mathrm{nrd}(\gamma)\,\mathrm{nrd}(\widehat{\alpha}) = \mathrm{nrd}(\widehat{\alpha}) \in F^\times_{>_\Omega 0}\backslash \widehat{F}^\times / \mathrm{nrd}(\widehat{O}^\times)$$

so the map is well-defined.

Similarly, the map (28.10.2) is a group homomorphism: an isomorphism $I \oplus I' \simeq J \oplus O$ gives $\mathrm{nrd}(\gamma\beta) = \mathrm{nrd}(\beta) = \mathrm{nrd}(\widehat{\alpha}\widehat{\alpha}')$. The map is surjective. If $[I]_{\mathsf{St}}$ is in the kernel and $\mathrm{nrd}(I)$ is trivial, then so too is $\mathrm{nrd}(\widehat{\alpha}_1)$ where $\widehat{\alpha}_1$ corresponds to $I \oplus O$; by strong approximation, this means that there exists $\beta \in \mathrm{GL}_2(B)$ and $\widehat{\mu} \in \mathrm{GL}_2(\widehat{O})$ such that $\widehat{\alpha}_1 = \beta\widehat{\mu}$ and so via β we have $I \oplus O \simeq O^{\oplus 2}$, which means $[I]_{\mathsf{St}} = [O]$. ◻

Exercises

Unless otherwise specified, let R be a global ring with eligible set S and $F = \operatorname{Frac} R$, and let B be a quaternion algebra over F, and let $O \subset B$ be an R-order.

1. Give another proof weak approximation for $\mathbb{Q}$, as follows.

   (a) Let $S \subseteq \operatorname{Pl}\mathbb{Q}$ be a finite, nonempty set of places. Show that, for each $v \in S$, there exists $y_v \in \mathbb{Q}^\times$ such that $|y_v|_v < 1$ and $|y_v|_{v'} > 1$ for all $v' \in S \smallsetminus \{v\}$. *[Hint: let $y_\infty := \prod_{p \in S} 1/p$ and to get y_p, multiply y_∞ by a power of p.]*

   (b) For the elements y_v constructed in (a), show that for all $v \in S$ we have $1/(1 + y_v^n) \to 1 \in \mathbb{Q}_v$ and $1/(1 + y_v^n) \to 0 \in \mathbb{Q}_{v'}$ for $v' \neq v$. [How does this relate to the proof of the CRT?]

(c) Prove weak approximation. *[Hint: given $x_v \in \mathbb{Q}_v$ for each $v \in S$, show that*

$$z_n := \sum_{v \in S} \frac{x_v}{1 + y_v^n} \to x_v \in \mathbb{Q}_v$$

as $n \to \infty$.]

2. Show that for all $N \geq 1$, the group $\mathrm{SL}_2(\mathbb{Z}/N\mathbb{Z})$ is generated by two elements of order N.
3. Let F be a field and let $n \in \mathbb{Z}_{\geq 2}$. Show that the elementary matrices (which differ from the identity matrix in exactly one off-diagonal place) generate $\mathrm{SL}_n(F)$ as a group. *[Hint: Argue by induction, and reduce a matrix to the identity by elementary row and column operations.]*
4. Let $n \geq 2$.

   (a) Let R be a Euclidean domain. Show that the elementary matrices generate $\mathrm{SL}_n(R)$. *[Hint: In view of Lemma 28.3.3, argue by induction.]*

   (b) Using (a), show that Proposition 28.3.5 and Corollary 28.3.6 hold for SL_n.

5. Let B be a central simple algebra over the global field F. We say B **satisfies the S-Eichler condition** if there exists a place $v \in S$ such that B_v is not a division algebra. Show that if F is a number field and S is the set of archimedean places of F, then B satisfies the S-Eichler condition if and only if B is *not* a totally definite quaternion algebra.
6. In this exercise, we provide details in the proof of Proposition 28.3.5, considering elementary matrices of rings that are not necessarily domains.
   Let A be a commutative ring (with 1), not necessarily a domain. We say A is **Euclidean** if there exists a function $N\colon A \to \mathbb{Z}_{\geq 0}$ such that for all $a, b \in A$ with $b \neq 0$, there exists $q, r \in A$ such that $a = qb + r$ and $N(r) < N(b)$.
   In the first parts, we suppose A is Euclidean with respect to N and show that the situation is quite analogous to the case when A is a domain.

   (a) Show that for all $b \in A$ we have $N(b) \geq N(0)$ with equality if and only if $b = 0$.
   (b) Let $m = \min\{N(b) : b \neq 0\}$. Show that if $N(b) = m$ then $b \in A^\times$.
   (c) Show that every ideal of A is principal. [We call A a **principal ideal ring**.]

   We now consider examples.

   (d) Let A be an Artinian local principal ideal ring with a unique maximal ideal $\mathfrak{m} = \pi A$. (For example, we may take $A = R/\mathfrak{p}^e$ where R is a Dedekind domain, $\mathfrak{p}$ is a nonzero prime ideal, and $e \in \mathbb{Z}_{\geq 0}$.) Show that for all $x \in A$ with $x \neq 0$, there exists $u \in A^\times$ and a *unique* $n \in \mathbb{Z}_{\geq 0}$ such that $x = u\pi^n$. Conclude that A is Euclidean with $N(x) = n$.
   (e) If $A \simeq \prod_{i=1}^r A_i$ with each A_i an Artinian local principal ideal ring, show that A is Euclidean under $N(x) = \sum_{i=1}^r N_i(\pi_i(x))$ where N_i is as given in (a) for A_i and $\pi_i\colon A \to A_i$ is the projection.

   We conclude with the application.

(f) Let A be a Euclidean ring. Show that $\mathrm{SL}_2(A)$ is generated by elementary matrices. *[Hint: Show that the proof in Lemma 28.3.3 carries over.]*

7. Consider the quaternion algebra $B := (-11, -17 \mid \mathbb{Q})$ and let $O = \mathbb{Z} \oplus \mathbb{Z}i \oplus \mathbb{Z}j \oplus \mathbb{Z}ij$. Then B is definite, so strong approximation does not apply. Indeed, show that $\widehat{B}^1 \not\subseteq B^\times \widehat{O}^\times$ as follows.

   (a) Find $a, b, c, d, m \in \mathbb{Z}$ with $3 \nmid m$ such that
   $$a^2 + 11b^2 = c^2 + 17d^2 = 3m.$$

   (b) Now let $\widehat{\alpha} = (\alpha_p)_p \in \widehat{B}^\times$ be such that
   $$\alpha_3 = (a + bi)(c + dj)^{-1} = \frac{(a + bi)(c - dj)}{3m}$$
   and $\alpha_p = 1$ if $p \neq 3$. Show that $\widehat{\alpha} \in \widehat{B}^1$ and $\widehat{\alpha} \notin B^\times \widehat{O}^\times$. *[Hint: Observe that* $\mathrm{nrd}|_O$ *only represents* 9 *by* ± 3.*]*

   (c) Prove that the right O-ideals $I_1 := 3O + (a + bi)O$ and $I_2 := 3O + (c + di)O$ are not principal. How does this relate to (b)?

8. Let F be a number field with ring of integers R. Show that there is a finite set S of (rational) primes such that every totally positive element of R can be written as a sum of four squares of elements of F whose denominator is a product of primes in S. (We may not be able to write every such element as sum of four squares from R, but we only need denominators in S.)

9. Suppose that B is S-indefinite. Suppose O is locally norm-maximal. Give a direct proof using strong approximation that if $O \subseteq B$ is an R-order and I is an invertible right fractional O-ideal, then I is principal if and only if $[\mathrm{nrd}(I)]$ is trivial in $\mathrm{Cl}_\Omega R$. *[Hint: If $\alpha \in B^\times$ satisfies* $\mathrm{nrd}(\alpha)R = \mathrm{nrd}(I)$*, consider $\alpha^{-1}I$.]*

▹ 10. Show that $B^\times \widehat{O}^\times \cap \widehat{B}^1 = B^1 \widehat{O}^1$ if and only if $\mathrm{nrd}(O^\times) = F^\times_{>_\Omega 0} \cap \mathrm{nrd}(\widehat{O}^\times)$ (see 28.4.7).

11. Suppose that B is S-indefinite. Show that $\#\,\mathrm{Typ}\, O$ is a power of 2.

12. Suppose $\operatorname{disc} B = \mathfrak{D}$ and O is an Eichler order of level $\mathfrak{M}$ and reduced discriminant $\mathfrak{N} = \mathfrak{D}\mathfrak{M}$. Show that $\mathrm{Cl}_{GN(O)} R$ is the quotient of $(\mathrm{Cl}\, R)/(\mathrm{Cl}\, R)^2$ by the subgroup generated by the classes of ideals $\mathfrak{p} \mid \mathfrak{D}$ together with $\mathfrak{q} \mid \mathfrak{N}$ such that $\mathrm{ord}_{\mathfrak{q}}(\mathfrak{N})$ is odd.

    Conclude that for $B = \mathrm{M}_2(F)$ and $O = \begin{pmatrix} R & R \\ \mathfrak{N} & R \end{pmatrix}$, the type set $\mathrm{Typ}\, O$ is represented by the isomorphism classes of the orders $\begin{pmatrix} R & \mathfrak{b} \\ \mathfrak{N}\mathfrak{b}^{-1} & R \end{pmatrix}$ for $[\mathfrak{b}] \in \mathrm{Cl}_{GN(O)} R$.

This chapter is licensed under the terms of the Creative Commons Attribution-NonCommercial 4.0 International License (http://creativecommons.org/licenses/by-nc/4.0/), which permits any noncommercial use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license and indicate if changes were made.

The images or other third party material in this chapter are included in the chapter's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the chapter's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.

Chapter 29

Idelic zeta functions

In this chapter, we present an idelic formulation of zeta functions associated to central simple algebras over global fields and we prove that they have analytic continuation and functional equation.

To read beyond the first two introductory sections, the additional prerequisite of real analysis at the level of measure theory is recommended; happily, the statements of the essential conclusions (for example, the functional equation of the zeta function of a quaternion algebra) can be understood without this background.

29.1 ▹ Poisson summation and the Riemann zeta function

We begin with some essential motivation. Recall that the Riemann zeta function $\zeta(s) := \sum_{n=1}^{\infty} n^{-s}$ defined in (25.2.2). Following Riemann, we complete to the function $\xi(s) := \zeta(s)\Gamma_{\mathbb{R}}(s)$, where

$$\Gamma_{\mathbb{R}}(s) := \pi^{-s/2}\Gamma(s/2) \tag{29.1.1}$$

and $\Gamma(s)$ is the complex Γ-function (Exercise 26.2). Then $\xi(s)$ extends to a meromorphic function on $\mathbb{C}$ and satisfies the functional equation

$$\xi(1-s) = \xi(s). \tag{29.1.2}$$

(More generally, we saw in 26.8.2 that the Dedekind zeta function of a number field can be completed in an analogous manner, again with functional equation.)

In this introductory section, we sketch a proof of the functional equation and see it as a consequence of Poisson summation (arising naturally in Fourier analysis, still following Riemann). Then, following Tate we reinterpret this extra factor in a manner that realizes the zeta function as a zeta *integral* on an adelic space, giving a uniform description and making the whole setup more suitable for analysis.

To begin, we the function $\xi(s)$ itself as an integral. Looking at one term in $\xi(s)$, we have

$$\pi^{-s/2}\Gamma(s/2)n^{-s} = \pi^{-s/2}n^{-s}\int_0^{\infty} e^{-x}x^{s/2}\frac{\mathrm{d}x}{x} = \int_0^{\infty} e^{-\pi n^2 u}u^{s/2}\frac{\mathrm{d}u}{u}. \tag{29.1.3}$$

© The Author(s) 2021

J. Voight, *Quaternion Algebras*, Graduate Texts in Mathematics 288,
https://doi.org/10.1007/978-3-030-56694-4_29

Accordingly, we define the theta function

$$\Theta(u) := \sum_{n=-\infty}^{\infty} e^{-\pi n^2 u} = 1 + 2\sum_{n=1}^{\infty} e^{-\pi n^2 u} \tag{29.1.4}$$

studied by Jacobi, convergent absolutely to a holomorphic function on the right half-plane $\operatorname{Re} u > 0$. Summing over $n \geq 1$ gives

$$\xi(s) = \frac{1}{2}\int_0^\infty (\Theta(u) - 1)u^{s/2}\frac{\mathrm{d}u}{u} \tag{29.1.5}$$

valid for $\operatorname{Re} s > 1$, where the integral converges (so we may justify interchanging summation and integration). We will soon rewrite this integral so as to extend its definition to all $s \in \mathbb{C}$ apart from $s = 0, 1$.

Let $f: \mathbb{R} \to \mathbb{C}$ be a **Schwartz** function, so f is infinitely differentiable and every derivative **decays rapidly** (for all $m, n \geq 0$ we have $|f^{(m)}(x)| = O(x^{-n})$ as $|x| \to \infty$). We define the **Fourier transform** $f^\vee : \mathbb{R} \to \mathbb{C}$ of f by

$$f^\vee(y) = \int_{-\infty}^{\infty} e^{-2\pi i x y} f(x)\,\mathrm{d}x. \tag{29.1.6}$$

Theorem 29.1.7 (Poisson summation). *We have*

$$\sum_{m=-\infty}^{\infty} f(m) = \sum_{n=-\infty}^{\infty} f^\vee(n), \tag{29.1.8}$$

with both sums converging absolutely.

Proof. The condition that f is Schwartz ensures good analytic behavior allowing the interchange of sum and integral, the details of which we elide. Let $g(x) := \sum_{m=-\infty}^{\infty} f(x+m)$. Then g is periodic with period 1, so by Fourier expansion we have $g(x) = \sum_{n=-\infty}^{\infty} a_n e^{2\pi i n x}$ where

$$\begin{aligned} a_n &:= \int_0^1 g(x)e^{-2\pi i n x}\,\mathrm{d}x = \int_0^1 \sum_{m=-\infty}^{\infty} f(x+m)e^{-2\pi i n x}\,\mathrm{d}x \\ &= \sum_{m=-\infty}^{\infty} \int_0^1 f(x+m)e^{-2\pi i n(x+m)}\,\mathrm{d}x = \int_{-\infty}^{\infty} f(x)e^{-2\pi i n x}\,\mathrm{d}x = f^\vee(n). \end{aligned}$$

Thus

$$\sum_{m=-\infty}^{\infty} f(m) = g(0) = \sum_{n=-\infty}^{\infty} f^\vee(n). \qquad \square$$

Now take $f(x) = e^{-\pi x^2}$. Then f is Schwartz. By contour integration and using $\int_{-\infty}^{\infty} e^{-\pi x^2}\,\mathrm{d}x = 1$, we conclude that $f^\vee(y) = f(y)$ for all y. For $u > 0$, let $f_u(x) := f(ux)$; then $f_u^\vee(y) = u^{-1} f_{1/u}(y)$ by change of variable. Applying Poisson summation to $f_{\sqrt{u}}(x)$ then gives

$$\Theta(u) = \frac{1}{\sqrt{u}}\Theta(1/u). \tag{29.1.9}$$

The equation (29.1.9) implies the functional equation (29.1.2) for $\xi(s)$ as follows: we split up the integral (29.1.5) as

$$\xi(s) = -\frac{1}{s} + \frac{1}{2}\int_0^1 \Theta(u)u^{s/2}\frac{\mathrm{d}u}{u} + \frac{1}{2}\int_1^\infty (\Theta(u)-1)u^{s/2}\frac{\mathrm{d}u}{u}$$

and apply the change of variable $u \leftarrow 1/u$ to obtain

$$\int_0^1 \Theta(u)u^{s/2}\frac{\mathrm{d}u}{u} = \int_1^\infty \Theta(u)u^{(1-s)/2}\frac{\mathrm{d}u}{u} = \frac{2}{s-1} + \int_1^\infty (\Theta(u)-1)u^{(1-s)/2}\frac{\mathrm{d}u}{u}.$$

Putting these together, we have

$$\xi(s) = \frac{1}{2}\int_1^\infty (\Theta(u)-1)(u^{s/2}+u^{(1-s)/2})\frac{\mathrm{d}u}{u} - \frac{1}{s} - \frac{1}{1-s} \tag{29.1.10}$$

which is sensible as a meromorphic function for all $s \in \mathbb{C}$, holomorphic except for $s = 0, 1$, with the right-hand side visibly invariant under $s \leftarrow 1-s$. This establishes (29.1.2).

This method extends to prove the functional equation for the L-series $L(s,\chi)$ where χ is a Dirichlet character (now involving a Gauss sum); Hecke extended this method (generalizing the appropriate theta functions) to prove the functional equation for a wider class of functions, including the Dedekind zeta functions.

29.2 ▹ Idelic zeta functions, after Tate

Now convinced of the utility of integral representations, we seek to put the finite places on an equal footing. In this way, the inclusion $\mathbb{Z} \subseteq \mathbb{R}$ in the Fourier analysis above is replaced by $\mathbb{Q} \subseteq \underline{\mathbb{Q}}$. Recall that $\zeta(s) = \prod_p \zeta_p(s)$ where

$$\zeta_p(s) := \sum_{e=0}^{\infty} p^{-es} = (1-p^{-s})^{-1};$$

we recover these factors from an integral. First we need a measure to integrate against. We have $\mathbb{Z}_p = \varprojlim_n \mathbb{Z}/p^n\mathbb{Z}$ as a projective limit with compatible projection maps $\pi_n : \mathbb{Z}_p \to \mathbb{Z}/p^n\mathbb{Z}$. We define the measure on $\mathbb{Z}_p$ as the projective limit of the counting measures on each $\mathbb{Z}/p^n\mathbb{Z}$ with total measure 1, i.e., for a set $E \subseteq \mathbb{Z}_p$ we define

$$\mu_p(E) := \lim_{n\to\infty} \frac{\#\pi_n(E)}{p^n}$$

when this limit exists. The measure extends additively to $\mathbb{Q}_p = \mathbb{Z}_p[1/p]$ and is invariant under additive translation

$$\mu_p(a+E) = \mu_p(E) \quad \text{for all } a \in \mathbb{Z}_p;$$

accordingly, μ_p is the *standard Haar measure* on $\mathbb{Q}_p$ (see 29.3). We have normalized the measure so that

$$\mu_p(\mathbb{Z}_p) = \int_{\mathbb{Z}_p} \mathrm{d}\mu_p(x) = 1.$$

Since $\mathbb{Z}_p^\times = \mathbb{Z}_p \smallsetminus p\mathbb{Z}_p$ and $\mathbb{Z}_p = \bigsqcup_{a=0}^{p-1}(a + p\mathbb{Z}_p)$ we have $\mu(p\mathbb{Z}_p) = 1/p$ and

$$\mu_p(\mathbb{Z}_p^\times) = 1 - 1/p. \tag{29.2.1}$$

Similarly, we have a *standard Haar measure* $\mu_p^\times$ on $\mathbb{Q}_p^\times$ by

$$\mathrm{d}\mu_p^\times(x) := \left(1 - \frac{1}{p}\right)^{-1} \frac{\mathrm{d}\mu_p(x)}{|x|_p};$$

the measure is invariant under $x \leftarrow ax$ for $a \in \mathbb{Q}_p^\times$ as well as under the substitution $x \leftarrow x^{-1}$, and with our normalization we have

$$\mu_p^\times(\mathbb{Z}_p^\times) = \left(1 - \frac{1}{p}\right)^{-1} \mu_p(\mathbb{Z}_p^\times) = 1.$$

For a complex number $s \in \mathbb{C}$, on our way to recover $\zeta_p(s)$ we consider the (Lebesgue) integral

$$\int_{\mathbb{Z}_p \smallsetminus \{0\}} |x|_p^s \, \mathrm{d}\mu_p^\times(x). \tag{29.2.2}$$

This looks a bit weird at first, because it is not over a subgroup of $\mathbb{Q}_p^\times$ or anything. Nevertheless, it works! For every nonzero $x \in \mathbb{Z}_p \smallsetminus \{0\}$, we may write $x = p^e x_0$ with $x_0 \in \mathbb{Z}_p^\times$ and $e \geq 0$, therefore the integral can be written as a sum over the level sets $p^e\mathbb{Z}_p^\times$:

$$\int_{\mathbb{Z}_p \smallsetminus \{0\}} |x|_p^s \, \mathrm{d}\mu_p^\times(x) = \sum_{e=0}^{\infty} p^{-es} \mu^\times(\mathbb{Z}_p^\times) = \left(1 - p^{-s}\right)^{-1} = \zeta_p(s). \tag{29.2.3}$$

It is more common to rewrite this as an integral over $\mathbb{Q}_p^\times$ by letting Ψ_p be the characteristic function of $\mathbb{Z}_p \smallsetminus \{0\}$, so that

$$\int_{\mathbb{Z}_p \smallsetminus \{0\}} |x|_p^s \, \mathrm{d}\mu_p^\times(x) = \int_{\mathbb{Q}_p^\times} |x|_p^s \Psi_p(x) \, \mathrm{d}\mu_p^\times(x). \tag{29.2.4}$$

In a similar fashion, we define the measure μ_∞ on $\mathbb{R}$ by $\mathrm{d}x/|x|$, and to match (29.1.1) we define $\Psi_\infty(x) = e^{-\pi x^2}$. Putting these together, on the idele group $\underline{\mathbb{Q}}^\times$ we define the product measure $\underline{\mu}^\times = \prod_v \mu_v^\times$, the function $\underline{\Psi}(\underline{x}) = \prod_v \Psi_v(x_v)$, and absolute value $\|\underline{x}\| = \prod_v |x_v|_v$ (trivial on $\mathbb{Q}^\times$ by the product formula). We have then repackaged the zeta function as an adelic integral

$$\xi(s) = \int_{\underline{\mathbb{Q}}^\times} |\underline{x}|^s \underline{\Psi}(\underline{x}) \, \mathrm{d}\underline{\mu}^\times(\underline{x}). \tag{29.2.5}$$

This is really nice!

Rewritten in this idelic way, Tate [Tate67] in his Ph.D. thesis elegantly proved the functional equation for a wide class of zeta functions (and L-functions):

> [T]he role of Hecke's complicated theta-formulas for theta functions formed over a lattice in the n-dimensional space of classical number theory can be played by a simple Poisson formula [...], the number theoretic analogue of the Riemann–Roch theorem" [Tate67, p. 305–306].

In this chapter, we use this method of Poisson summation and idelic integrals to prove the basic properties of the zeta function of a central simple algebra over a global field (Main Theorem 29.10.1). Translated back into classical language, we prove as a consequence the key result (the crux of which is Theorem 26.8.19) announced in section 26.8.

Main Theorem 29.2.6. *Let F be a number field with ring of integers $R = \mathbb{Z}_F$ and let B be a quaternion algebra over F with maximal order $\mathcal{O}$. Let*

$$\zeta_B(s) := \sum_{I \subseteq \mathcal{O}} \mathsf{N}(I)^{-s}$$

be the sum over nonzero right $\mathcal{O}$-ideals, where N is the absolute (counting) norm (16.4.7) and let $\xi_B(s)$ be its completion (26.8.13). Then $\xi_B(s)$ has meromorphic continuation to $\mathbb{C}$, holomorphic away from $\{0, 1/2, 1\}$ with simple poles at $s = 0, 1$, and it satisfies the functional equation

$$\xi_B(1-s) = \xi_B(s). \tag{29.2.7}$$

Moreover, if B is a division algebra, then $\xi_B(s)$ is holomorphic at $s = 1/2$.

Main Theorem 29.2.6 is an *analytic* result with key *arithmetic* consequences, including the classification of quaternion algebras over global fields, as we saw in section 26.8. More than that, the evaluation of the residue will further give rise to a volume formula (Main Theorem 39.1.8) that generalizes the Eichler mass formula (as in the proof of Proposition 26.5.10).

The developments in this chapter have a rich history, and they generalize *vastly* beyond this text: see Remark 29.10.24. We restrict ourselves to the case of zeta functions both because this suffices for our main applications and because it is a good stepping stone to the more general theory. Although this chapter is quite technical, the reader's forbearance will ultimately be rewarded!

29.3 Measures

In this section, we define the local measures we will use. As references for this section and the next, see Bekka–de la Harpe–Valette [BHV2008, Appendices A and B], Deitmar [Dei2005], Deitmar–Echterhoff [DE2009, Chapters 1 and 3, Appendix B], Loomis [Loo53], and Ramakrishnan–Valenza [RM99, Chapters 1–4] as references on harmonic analysis, and Vignéras [Vig80a, §II.4] and Weil [Weil74, Chapter XI] for the present context.

Let G be a Hausdorff, locally compact, second countable topological group. For example, we may take $G = \mathrm{SL}_2(\mathbb{R})$ or $G = \mathrm{SL}_n(\mathbb{R})$ (or more generally a semisimple real Lie group). A **Borel measure** on G is a countably additive function, with values

in $[0, \infty]$, defined on the σ-algebra generated by open sets in G under complement and finite or countable unions.

Definition 29.3.1. A **Radon measure** on G is a Borel measure that is finite on compact sets. A **left Haar measure** on G is a nonzero Radon measure μ that is left **translation-invariant**, so $\mu(gE) = \mu(E)$ for all Borel subsets $E \subseteq G$ and all $g \in G$.

The notion of *right Haar measure* is defined similarly.

Remark 29.3.2. Since G is second countable and locally compact, every open subset of G is σ**-compact** (the union of a countable collection of compact subspaces), so a Radon measure is necessarily **regular**, i.e., the measure of a Borel set is the infimum of the measures of its open supersets and the measure of an open set is the supremum of the measures of its compact subsets: see Rudin [Rud87, Theorem 2.18].

Proposition 29.3.3. *G admits a left Haar measure that is unique up to scaling by an element of* $\mathbb{R}_{>0}$.

Proof. See e.g. Deitmar–Echterhoff [DE2009, Theorem 1.3.4] or Diestel–Spalsbury [DS2014, §7.2]. □

We will construct Haar measures explicitly as we need them, so we need not appeal to the general result of Proposition 29.3.3 beyond the uniqueness statement which is itself straightforward to establish: see Exercise 29.4.

Example 29.3.4. On $G = \mathbb{R}^n$ under addition, a left (and right) Haar measure is given by the usual Lebesgue measure.

Example 29.3.5. Suppose G has the discrete topology (all sets are open). Then the counting measure $\mu(E) = \#E \in \mathbb{Z}_{\geq 0} \cup \{\infty\}$ is a left (and right) Haar measure.

From now on, let μ be a left Haar measure on G.

29.3.6. In general, a left Haar measure need not also be right translation-invariant. For $g \in G$, the measure defined by $\mu_g(E) := \mu(Eg)$ for a Borel set E is again a left Haar measure, so by Proposition 29.3.3, we have $\mu_g = \Delta_G(g)\mu$ for some $\Delta_G(g) \in \mathbb{R}_{>0}$. Since μ is unique up to scaling, the function $\Delta_G(g)$ does not depend on μ.

Definition 29.3.7. The map $\Delta_G \colon G \to \mathbb{R}_{>0}$ is the **modular function** of G. We say G is **unimodular** if Δ_G is identically 1.

By 29.3.6, G is unimodular if and only if every left Haar measure is a right Haar measure.

Lemma 29.3.8. *The modular function Δ_G is a homomorphism.*

Proof. Exercise 29.5. □

Example 29.3.9. If G is abelian or discrete (see Example 29.3.5), then G is unimodular: every left Haar measure is a right Haar measure. (There are many other important classes of groups that are unimodular, including discrete (e.g., finite) groups, compact groups, and the others we will in what follows.)

29.3.10. In this paragraph, we briefly review the theory of integration we need; for further details, see Deitmar–Echterhoff [DE2009, Appendix B.1]. Let $f\colon G \to \mathbb{C}$ be a complex-valued function on G. We say f is **measurable** if for all Borel sets $E \subseteq \mathbb{C}$, the subset $f^{-1}(E)$ is measurable; we say f is **real(-valued)** if $f(G) \subseteq \mathbb{R}$ and is **nonnegative** if $f(G) \subseteq \mathbb{R}_{\geq 0}$.

If E is a measurable set then the characteristic function 1_E of E (equal to 1 on E and 0 outside E) is defined to have integral

$$\int_G 1_E(x)\,\mathrm{d}\mu(x) := \mu(E) = \int_E \mathrm{d}\mu(x). \tag{29.3.11}$$

A **step function** is a finite $\mathbb{C}$-linear combination of characteristic functions of measurable sets, and we define the integral of a step function by linearity using (29.3.11).

If f is measurable and nonnegative, we define

$$\int_G f(x)\,\mathrm{d}\mu(x) := \sup\left\{\int_G g(x)\,\mathrm{d}\mu(x) : \begin{array}{l} g \text{ a nonnegative step function such} \\ \text{that } g(x) \leq f(x) \text{ for all } x \in G \end{array}\right\}$$

and say f is **integrable** if $\int_G f(x)\,\mathrm{d}\mu(x) < \infty$.

If f is real we say f is **integrable** if both $f^+ := \max(f, 0)$ and $f^- := -\min(f, 0)$ are integrable. If f is real and integral, we then define

$$\int_G f(x)\,\mathrm{d}\mu(x) := \int_G f^+(x)\,\mathrm{d}\mu(x) - \int_G f^-(x)\,\mathrm{d}\mu(x). \tag{29.3.12}$$

Finally, a general f is **integrable** if $\operatorname{Re} f$ and $\operatorname{Im} f$ are integrable, in which case

$$\int_G f(x)\,\mathrm{d}\mu(x) := \int_G \operatorname{Re} f(x)\,\mathrm{d}\mu(x) + i\int_G \operatorname{Im} f(x)\,\mathrm{d}\mu(x).$$

29.3.13. Let H be another Hausdorff, locally compact, second countable topological group, and let $\phi\colon G \to H$ be continuous surjection with kernel $N := \ker\phi$. Then we have an exact sequence

$$1 \to N \to G \xrightarrow{\phi} H \to 1.$$

We say that left Haar measures on G, N, H are **compatible** if for all integrable functions $f\colon G \to \mathbb{C}$,

$$\int_G f(x)\,\mathrm{d}\mu(x) = \int_H \left(\int_N f(zy)\,\mathrm{d}\mu(z)\right)\mathrm{d}\mu(\phi(y)). \tag{29.3.14}$$

Given measures on two terms, there exists a unique compatible measure on the third—but note, this measure depends on the exact sequence (Exercise 29.7).

29.4 Modulus and Fourier inversion

We continue our background review with notation from the previous section; we now treat the modulus of an automorphism and present Fourier inversion.

29.4.1. Let $\phi \in \mathrm{Aut}(G)$ be a continuous automorphism. Then the measure defined by $\mu_\phi(E) := \mu(\phi(E))$ is again a left Haar measure. By Proposition 29.3.3, there exists a unique $\|\phi\| \in \mathbb{R}_{>0}$ such that

$$\mu(\phi E) = \|\phi\|\mu(E) \tag{29.4.2}$$

for all measurable sets E. In particular, for all integrable functions f on G we have

$$\int_G f(x)\,\mathrm{d}\mu(x) = \int_G f(\phi(x))\,\mathrm{d}\mu(\phi(x)) = \|\phi\| \int_G f(\phi(x))\,\mathrm{d}\mu(x). \tag{29.4.3}$$

Definition 29.4.4. The **modulus** of $\phi \in \mathrm{Aut}(G)$ is $\|\phi\| \in \mathbb{R}_{>0}$ as defined in (29.4.2).

The definition of the modulus is independent of the choice of Haar measure μ, since Haar measure is unique up to scaling and such a scalar cancels in (29.4.2).

Example 29.4.5. Let $g \in G$; then the conjugation automorphism $\phi_g(x) = g^{-1}xg$ has $\|\phi_g\| = \Delta_G(g)$, because μ is left-invariant. In particular, G is unimodular if and only if all conjugation maps have trivial modulus.

Now let A be a Hausdorff, locally compact, second countable topological *ring* and let μ be a left Haar measure on the additive group of A. Since A is abelian, A is unimodular and μ is also a right Haar measure.

29.4.6. Let $a \in A^\times$. Then the left multiplication map $\lambda_a : A \to A$ by $x \mapsto ax$ is a continuous automorphism of A as an additive abelian group, so we may define its **modulus** $\|a\| := \|\lambda_a\|$ by (29.4.2): symbolically, we write

$$\|a\| = \frac{\mathrm{d}\mu(ax)}{\mathrm{d}\mu(x)}. \tag{29.4.7}$$

We have $\|ab\| = \|a\|\|b\|$ for all $a, b \in A^\times$ (Exercise 29.6).

29.4.8. The measure on $A^\times$ defined by

$$\mathrm{d}\mu^\times(x) := \frac{\mathrm{d}\mu(x)}{\|x\|} \tag{29.4.9}$$

is a (multiplicative) Haar measure by 29.4.6.

We conclude this introductory section with the Fourier inversion formula.

29.4.10. Let $\mathbb{C}^1 := \{z \in \mathbb{C} : |z| = 1\}$ be the circle group. A **unitary character** of A is a continuous homomorphism $\chi\colon A \to \mathbb{C}^1$, considering A as an additive group. Let $A^\vee := \mathrm{Hom}(A, \mathbb{C}^1)$ be the **unitary character group** of A under pointwise multiplication, and equip $A^\vee$ with the compact-open topology (as a closed subset of the set of all continuous maps $A \to \mathbb{C}^1$).

Remark 29.4.11. We reserve the term *character* for continuous group homomorphisms $A \to K^\times$, where K is a field (of values for the character); this notion makes sense for any field K. Some authors call unitary characters just *characters*, then calling our characters instead *quasi-characters*.

Now *suppose* that there exists $\psi \in A^\vee$ such that the map

$$\begin{aligned} A &\to A^\vee \\ x &\mapsto (y \mapsto \psi(xy)) \end{aligned}$$

is an isomorphism of topological groups. (This is a hypothesis on A, and may depend on a choice; in the cases we consider below, we will identify a standard such ψ.)

Definition 29.4.12. For $f\colon A \to \mathbb{C}$ continuous and integrable, the **Fourier transform** of f (relative to ψ, μ) is the function

$$\begin{aligned} f^\vee &\colon A \to \mathbb{C} \\ f^\vee(x) &= \int_A f(y)\psi(xy)\,\mathrm{d}\mu(y). \end{aligned}$$

Theorem 29.4.13 (Fourier inversion). *There exists a unique Haar measure τ on A (depending on ψ) such that for all $f\colon A \to \mathbb{C}$ continuous and integrable with $f^\vee$ (defined relative to τ) continuous and integrable, we have*

$$f(x) = \int_A f^\vee(y)\overline{\psi(xy)}\,\mathrm{d}\tau(y). \tag{29.4.14}$$

Proof. The proof of this theorem is beyond the scope of this textbook, and we use it as a black box; see e.g. Deitmar–Echterhoff [DE2009, Theorem 3.5.8] or Folland [Fol95, Theorems 4.32–4.33]. □

The normalized measure τ in Theorem 29.4.13 is called the **self-dual measure** on A (with respect to ψ).

29.5 Local measures and zeta functions: archimedean case

Let B be a finite-dimensional simple algebra over the local field $F = \mathbb{R}$. A good general reference for the next three sections is Weil [Weil82, Chapter II].

29.5.1. The (additive) Haar measure μ on B is the usual (Lebesgue) measure, normalized as follows: letting $n = \dim_{\mathbb{R}} B$, we choose an $\mathbb{R}$-basis $e_1, \dots, e_n$ for B, so that we may write $x = \sum_i x_i e_i \in B$ with x_i are coordinates on B, and we define

$$\mathrm{d}x := |d(e_1, \dots, e_j)|^{1/2}\,\mathrm{d}x_1 \cdots \mathrm{d}x_n$$

where d is the discriminant defined by (15.2.1) and the reduced trace is taken on B as an $\mathbb{R}$-algebra. By Lemma 15.2.5, we see that this measure is independent of the choice of basis e_i.

Another application of Lemma 15.2.5 then gives the modulus

$$\|\alpha\| = |\mathrm{Nm}_{B|\mathbb{R}}(\alpha)| \tag{29.5.2}$$

for all $\alpha \in B^\times$.

Example 29.5.3. We compute:

$$
\begin{aligned}
\mathrm{d}x &= \mathrm{d}x, && \text{for } x \in \mathbb{R};\\
\mathrm{d}z &= 2\,\mathrm{d}x\,\mathrm{d}y, && \text{for } z = x + y\sqrt{-1} \in \mathbb{C};\\
\mathrm{d}\alpha &= 4\,\mathrm{d}t\,\mathrm{d}x\,\mathrm{d}y\,\mathrm{d}z, && \text{for } \alpha = t + xi + yj + zk \in \mathbb{H};\\
\mathrm{d}\alpha &= \prod_{i,j} \mathrm{d}x_{ij}, && \text{for } \alpha = (x_{ij})_{i,j} \in \mathrm{M}_n(\mathbb{R});\\
\mathrm{d}\alpha &= 2^{n^2} \prod_{i,j} \mathrm{d}x_{ij}\,\mathrm{d}y_{ij}, && \text{for } \alpha = (x_{ij} + y_{ij}\sqrt{-1})_{i,j} \in \mathrm{M}_n(\mathbb{C}).
\end{aligned}
$$

And:

$$
\begin{aligned}
\|x\| &= |x|, && \text{for } x \in \mathbb{R}^\times;\\
\|z\| &= |z|^2, && \text{for } z \in \mathbb{C}^\times;\\
\|\alpha\| &= \mathrm{nrd}(\alpha)^2, && \text{for } \alpha \in \mathbb{H}^\times;\\
\|\alpha\| &= |\det(\alpha)|^n, && \text{for } \alpha \in \mathrm{GL}_n(\mathbb{R});\\
\|\alpha\| &= |\det(\alpha)|^{2n}, && \text{for } \alpha \in \mathrm{GL}_n(\mathbb{C}).
\end{aligned}
$$

The modulus for $\mathbb{R}$ is the usual absolute value, whereas the modulus for $\mathbb{C}$ is the *square* of the absolute value (cf. Remark 12.2.3), explaining conventions on normalized absolute values in the product formula 14.4.6.

29.5.4. As in (29.4.9), the (multiplicative) Haar measure $\mu^\times$ on $B^\times$ is defined by

$$\mathrm{d}\mu^\times(\alpha) := \frac{\mathrm{d}\mu(\alpha)}{\|\alpha\|}. \tag{29.5.5}$$

29.5.6. We define the **standard unitary character**

$$
\begin{aligned}
\psi\colon B &\to \mathbb{C}^1\\
\alpha &\mapsto \exp(-2\pi i\,\mathrm{trd}(\alpha))
\end{aligned}
$$

with reduced trace taken on B as an $\mathbb{R}$-algebra. We note that $\psi(\alpha\beta) = \psi(\beta\alpha)$ for all $\alpha, \beta \in B$.

A fundamental result in standard Fourier analysis (generalizing the case $B = \mathbb{R}, \mathbb{C}$ and following from it in the same way) is the following proposition.

Proposition 29.5.7. *The standard unitary character induces an isomorphism $B \xrightarrow{\sim} B^\vee$ of topological groups, and the measure μ defined in 29.5.1 is self-dual (with respect to ψ).*

In light of Proposition 29.5.7, we will also write $\tau = \mu$ for the measure defined above.

29.5.8. The exact sequence

$$1 \to \mathbb{H}^1 \to \mathbb{H}^\times \xrightarrow{\mathrm{nrd}} \mathbb{R}_{>0} \to 1$$

defines a measure τ^1 on $\mathbb{H}^1$ by compatibility, taking the normalized measures $\tau^\times$ on $\mathbb{H}^\times$ and on $\mathbb{R}_{>0} \leq \mathbb{R}^\times$.

Lemma 29.5.9. *We have* $\tau^1(\mathbb{H}^1) = 4\pi^2$.

Proof. Let $\rho > 0$ and let

$$E = \{\alpha \in \mathbb{H}^\times : \mathrm{nrd}(\alpha) \leq \rho^2\}$$

be the punctured ball of radius ρ. Let f be the function $\|\alpha\|$ on E, zero elsewhere. On the one hand,

$$\int_{\mathbb{H}^\times} f(\alpha)\,\mathrm{d}\tau^\times(\alpha) = \int_E \|\alpha\| \frac{\mathrm{d}\tau(\alpha)}{\|\alpha\|} = \tau(E).$$

Recalling that τ is 4 times Lebesgue measure, and that the Lebesgue measure of a sphere of radius ρ has volume $\pi^2\rho^4/2$, we get $\tau(E) = 2\pi^2\rho^4$. On the other hand, by compatibility, this integral is equal to

$$\begin{aligned}\int_{\mathbb{R}_{>0}} \int_{\mathbb{H}^1} f(\alpha_1 r)\,\mathrm{d}\tau^1(\alpha_1)\,\mathrm{d}^\times(r^2) &= \int_0^\rho \int_{\mathbb{H}^1} \|\alpha_1 r\| \mathrm{d}\tau^1(\alpha_1)\,\frac{2r\,\mathrm{d}r}{r^2} \\ &= 2\tau^1(\mathbb{H}^1)\int_0^\rho r^3\,\mathrm{d}r = \tau^1(\mathbb{H}^1)\frac{\rho^4}{2}.\end{aligned}$$

We conclude that $\tau^1(\mathbb{H}^1) = 4\pi^2$. □

We will need further functions to integrate, so we make the following definition. When we consider functions on B, we may think of choosing an $\mathbb{R}$-basis $e_1, \dots, e_n$ of B, identifying $B \simeq \mathbb{R}^n$ and writing $x = \sum_i x_i e_i \in B$; our definitions will be independent of this choice of basis.

Definition 29.5.10. A function $f\colon B \to \mathbb{C}$ **decays rapidly** if for all $n \geq 0$ we have $|f(x)| = O((\max_i |x_i|)^{-n})$ as $\max_i |x_i| \to \infty$.

A function $f\colon B \to \mathbb{C}$ is **Schwartz(–Bruhat)** if f is infinitely differentiable and every partial derivative of f decays rapidly.

29.5.11. Let ${}^*\colon B \to B$ be the conjugate transpose involution (see 8.4.3) on B. Let $Q(\alpha) := \mathrm{trd}(\alpha\alpha^*)$ for $\alpha \in B$. We compute that

$$\begin{aligned} Q(x) &= x^2, && \text{for } x \in \mathbb{R};\\ Q(z) &= 2|z|^2 = 2(x^2+y^2), && \text{for } z = x + y\sqrt{-1} \in \mathbb{C};\\ Q(\alpha) &= 2\,\mathrm{nrd}(\alpha), && \text{for } \alpha \in \mathbb{H};\\ Q(\alpha) &= \sum_{i,j} x_{ij}^2, && \text{for } \alpha = (x_{ij})_{i,j} \in \mathrm{M}_n(\mathbb{R}); \text{ and}\\ Q(\alpha) &= 2\sum_{i,j} |z_{ij}|^2, && \text{for } \alpha = (z_{ij})_{i,j} \in \mathrm{M}_n(\mathbb{C}). \end{aligned} \tag{29.5.12}$$

Definition 29.5.13. The **standard function** on B is

$$\begin{aligned}\Psi\colon B &\to \mathbb{C}\\ \Psi(\alpha) &= \exp(-\pi Q(\alpha)).\end{aligned} \tag{29.5.14}$$

It is straightforward to verify that the standard function is Schwartz.

Definition 29.5.15. For a Schwartz function Φ, we define the **(local) zeta function**

$$Z_B^{\Phi}(s) := \int_{B^\times} \Phi(\alpha)\|\alpha\|^s \,\mathrm{d}\tau^\times(\alpha) \tag{29.5.16}$$

wherever this integral converges. We abbreviate $Z_B^{\Psi}(s) = Z_B(s)$ for $\Phi = \Psi$ the standard function.

As in (26.8.3), we define

$$\begin{aligned}\Gamma_{\mathbb{R}}(s) &:= \pi^{-s/2}\Gamma(s/2)\\ \Gamma_{\mathbb{C}}(s) &:= 2(2\pi)^{-s}\Gamma(s).\end{aligned} \tag{29.5.17}$$

Lemma 29.5.18. *We have*

$$\begin{aligned}Z_{\mathbb{R}}(s) &= \Gamma_{\mathbb{R}}(s),\\ Z_{\mathbb{C}}(s) &= \pi\Gamma_{\mathbb{C}}(s),\\ Z_{\mathbb{H}}(s) &= \pi(2s-1)\Gamma_{\mathbb{R}}(2s)\Gamma_{\mathbb{R}}(2s-1) = 2\pi^2\Gamma_{\mathbb{R}}(2s)\Gamma_{\mathbb{R}}(2s+1),\\ Z_{\mathrm{M}_2(\mathbb{R})}(s) &= \pi\Gamma_{\mathbb{R}}(2s)\Gamma_{\mathbb{R}}(2s-1),\\ Z_{\mathrm{M}_2(\mathbb{C})}(s) &= 2\pi^3\Gamma_{\mathbb{C}}(2s)\Gamma_{\mathbb{C}}(2s-1).\end{aligned}$$

Proof. We have

$$Z_{\mathbb{R}}(s) = 2\int_0^\infty x^s e^{-\pi x^2}\,\frac{\mathrm{d}x}{x};$$

making the substitution $x \leftarrow \pi x^2$ gives the result. A similar argument with polar coordinates gives $Z_{\mathbb{C}}(s)$. The remaining integrals are pretty fun, so they are left as Exercise 29.8. ∎

Lemma 29.5.18 explains the provenance of the definitions of $\Gamma_{\mathbb{R}}(s), \Gamma_{\mathbb{C}}(s)$ from (26.8.3), and ultimately their appearance in (26.8.14). (A quick check on the constant in front is provided by $Z_B(1) = 1$; these particular choices of constants follow convention.) By comparison, Lemma 29.5.18 then shows that in general $Z_B^{\Phi}(s)$ has meromorphic continuation to $\mathbb{C}$.

29.6 Local measures: commutative nonarchimedean case

Now let F be a *non*archimedean local field, with valuation v, valuation ring R, and maximal ideal $\mathfrak{p} = \pi R \subseteq R$. Let $q := \#R/\mathfrak{p}$.

We begin by defining a Haar measure normalized so that R has measure 1; then we extend this to find the normalization in which the measure is self-dual with respect to a standard unitary character.

29.6.1. We have $R \cong \varprojlim_n R/\mathfrak{p}^n$ with projection maps $\pi_n \colon R \to R/\mathfrak{p}^n$. We define the (additive) measure μ on R by

$$\mu(E) := \lim_{n\to\infty} \frac{\#\pi_n(E)}{q^n}$$

for a subset $E \subset R$, when this limit exists. It is straightforward to check that E defines a Haar measure on R, and the measure extends to F by additivity. In this normalization, we have $\mu(R) = 1$, and more generally $\mu(a + \mathfrak{p}^n) = 1/q^n$ for $a \in R$ and $n \geq 0$.

Let $|\ |$ be the preferred absolute value (see 14.4.12), with $|\pi| = 1/q$.

Lemma 29.6.2. *For all $x \in F^\times$, we have*

$$\|x\| = |x|; \tag{29.6.3}$$

in particular, if $x \in R$, then $|x| = \mathsf{N}(Rx)^{-1}$ where N is the counting norm (16.4.7).

Proof. By multiplicativity, we may suppose $x \in R$. Then $[R : xR] = q^{v(x)} = \mathsf{N}(Rx)$. Adding up cosets of xR in R, we obtain

$$\mu(R) = [R : xR]\mu(xR) = q^{v(x)}\mu(xR);$$

so $\|x\| = \mu(xR)/\mu(R) = q^{-v(x)} = |x|$, as claimed. □

Example 29.6.4. Since $\mu(R^\times) = \mu(R) - \mu(\mathfrak{p})$ and

$$\mu(\mathfrak{p}) = \|\pi\|\mu(R) = 1/q,$$

we have

$$\mu(R^\times) = 1 - \frac{1}{q}. \tag{29.6.5}$$

29.6.6. We normalize the (multiplicative) Haar measure $\mu^\times$ on $F^\times$ by defining

$$\mathrm{d}\mu^\times(x) := \frac{1}{\mu(R^\times)} \frac{\mathrm{d}\mu(x)}{\|x\|} = (1 - 1/q)^{-1} \frac{\mathrm{d}\mu(x)}{\|x\|}$$

so that $\mu^\times(R^\times) = 1$.

Next, we consider the Fourier transform in this context.

29.6.7. We first define an additive homomorphism

$$\langle\ \rangle_F \colon F \to \mathbb{R}/\mathbb{Z}$$

as follows.

(a) If $F = \mathbb{Q}_p$, we define $\langle x\rangle_{\mathbb{Q}_p} \in \mathbb{Q}$ to be such that $0 \leq \langle x\rangle_{\mathbb{Q}_p} < 1$ and $x - \langle x\rangle_{\mathbb{Q}_p} \in \mathbb{Z}_p$.

(b) If $F = \mathbb{F}_q((t))$ and $x = \sum_i a_i t^i$ then we take $\langle x\rangle_F := \mathrm{Tr}_{\mathbb{F}_q/\mathbb{F}_p}(a_{-1})/p$.

(c) In general, if $F \supseteq F_0$ is a finite separable extension of fields, then we define

$$\langle x \rangle_F := \langle \operatorname{Tr}_{F/F_0} x \rangle_{F_0}.$$

Definition 29.6.8. The **standard unitary character** of F is

$$\begin{aligned} \psi_F \colon F &\to \mathbb{C}^1 \\ \psi_F(x) &= \exp(2\pi i \langle x \rangle_F). \end{aligned}$$

Proposition 29.6.9. *The standard unitary character ψ_F defines an isomorphism*

$$\begin{aligned} F &\xrightarrow{\sim} F^\vee \\ x &\mapsto (y \mapsto \psi(xy)) \end{aligned}$$

of topological groups.

Proof. Exercise 29.9. ☐

Proposition 29.6.10. *The following statements hold.*

(a) *If F is a local number field (so a finite extension of $\mathbb{Q}_p$ for a prime p), then the measure*

$$\tau := |\operatorname{disc}(R)|^{1/2}\mu = \mathsf{N}(\operatorname{disc} R)^{-1/2}\mu$$

is self-dual (with respect to ψ).

(b) *If F is a local function field, then μ is self-dual with respect to ψ.*

Proof. First consider (a), and suppose F is a finite extension of $\mathbb{Q}_p$. We seek to satisfy (29.4.14); the equation holds up to a constant, so we may choose appropriate f and x. We choose f as the characteristic function of R and $x = 0$, so that $f(0) = 1$ and

$$\int_F \left(\int_F f(z)\psi(yz)\,\mathrm{d}\mu(z) \right) \mathrm{d}\mu(y) = \int_F f(z) \left(\int_F \psi(yz)\,\mathrm{d}\mu(y) \right) \mathrm{d}\mu(z). \tag{29.6.11}$$

By character theory,

$$\int_F \psi(yz)\,\mathrm{d}\mu(y) = \mu(\{y \in F : \psi(yz) = 1\}) = \mu(\{y \in F : \operatorname{Tr}_{F/\mathbb{Q}_p}(yz) \in \mathbb{Z}_p\})$$

so (29.6.11) is equal to $\mu(R^\sharp)$ where

$$R^\sharp = \operatorname{codiff}(R) = \{x \in F : \operatorname{Tr}_{F/\mathbb{Q}_p}(xR) \in \mathbb{Z}_p\}.$$

Let x_i be a $\mathbb{Z}_p$-basis for R with $x_i^\sharp$ the dual basis, giving a $\mathbb{Z}_p$-basis for $R^\sharp$. By Lemma 15.6.17 we have

$$\operatorname{disc}(R) = [R^\sharp : R]_{\mathbb{Z}_p}$$

so since $\mu(R) = 1$ by additivity we have

$$\mu(R^\sharp) = |\operatorname{disc}(R)|^{-1}.$$

It follows then that

$$\tau = \mu(R^\sharp)^{-1/2}\mu = |\operatorname{disc}(R)|^{1/2}\mu$$

is self-dual. ☐

29.7 Local zeta functions: nonarchimedean case

Continuing with F a nonarchimedean local field, let B be a finite-dimensional simple (not necessarily central) algebra over F, with maximal order O, and $n^2 = \dim_{Z(B)} B$.

29.7.1. The (additive) Haar measure μ on B is defined as in 29.6.1 as a projective limit, normalized so that $\mu(O) = 1$.

We compute that $\|\alpha\|_B = \|\mathrm{Nm}_{B|F}(\alpha)\|_F$, so that if B is central over F with $\dim_F B = n^2$, then $\|\alpha\|_B = \|\mathrm{nrd}(\alpha)^n\|_F$. In the remainder of this section, we work over B and drop the subscript $_B$.

29.7.2. For the (multiplicative) Haar measure $\mu^\times$ on $B^\times$, we use the same normalization factor (29.6.5) as for $F^\times$, defining the **normalized** measure

$$d\mu^\times(\alpha) := (1 - 1/q)^{-1} \frac{d\mu(\alpha)}{\|\alpha\|}. \tag{29.7.3}$$

29.7.4. We extend the standard unitary character on F, defined in 29.6.7, to a **standard unitary character** on B by

$$\psi_B(\alpha) := \psi_F(\mathrm{trd}(\alpha))$$

for $\alpha \in B$. Again, $\psi(\alpha\beta) = \psi(\beta\alpha)$ for all $\alpha, \beta \in B$.

If F is a local number field containing $\mathbb{Q}_p$, let $R_0 = \mathbb{Z}_p$; if F is a local function field, let $R_0 = R$.

Definition 29.7.5. The **absolute discriminant** of B is

$$\mathsf{D}(B) := \mathsf{N}(\mathrm{disc}_{R_0}(O)) = |\mathrm{disc}_{R_0}(O)|^{-1} \in \mathbb{Z}_{>0}.$$

The absolute discriminant is well-defined, independent of the choice of maximal order.

Example 29.7.6. If B is a quaternion algebra over F, then

$$\mathsf{D}(B) = \mathsf{N}(\mathrm{disc}_{R_0}(O)) = \mathsf{N}(\mathrm{disc}_{R_0}(R))^4 \mathsf{N}(\mathrm{discrd}_R(O))^2.$$

Proposition 29.7.7. *The standard unitary character ψ defines an isomorphism $B \xrightarrow{\sim} B^\vee$ of topological groups, and the measure*

$$\tau := \mathsf{D}(B)^{-1/2}\mu \tag{29.7.8}$$

is self-dual with respect to ψ.

Proof. The same arguments as in Propositions 29.6.9 and 29.6.10 apply, with appropriate modifications. □

29.7.9. Having normalized the multiplicative measure $\mu^\times$ and seeing the relevant modification in (29.7.8), just as in the case of $B = F$ we define

$$\tau^\times := \mathsf{D}(B)^{-1/2}\mu^\times. \tag{29.7.10}$$

We may now define the local zeta function in the nonarchimedean context.

Definition 29.7.11. A function $f\colon B \to \mathbb{C}$ is **Schwartz–Bruhat** if f is locally constant (for every $\alpha \in B$, there exists an open neighborhood $U \ni x$ such that $f|_U$ is constant) with compact support. A **standard function** Ψ on B is the characteristic function of a maximal order.

Taking an open cover of the support, we see that every Schwartz–Bruhat function can be expressed as a (finite) $\mathbb{C}$-linear combination of characteristic functions of compact open subsets of B.

Definition 29.7.12. For a Schwartz–Bruhat function Φ, we define the **(local) zeta function**

$$Z_B^{\Phi}(s) := \int_{B^\times} \|\alpha\|^s \Phi(\alpha)\,\mathrm{d}\tau^\times(\alpha). \tag{29.7.13}$$

We write

$$Z_B(s) := Z_B^{\Psi}(s) = \int_{\mathcal{O}\cap B^\times} \|\alpha\|^s\,\mathrm{d}\tau^\times(\alpha) \tag{29.7.14}$$

for the zeta function with respect to a standard function Ψ; this is well-defined (independent of Ψ) as any two maximal orders are conjugate, and $\|\alpha\|$ is well-defined on conjugacy classes by 29.7.1.

We begin with a basic convergence estimate.

Lemma 29.7.15. *The function $Z_B^{\Phi}(s)$ converges for* $\operatorname{Re} s > 1$.

Proof. We reduce to the case where Φ is the characteristic function of a compact open set U, since Φ is a (finite) $\mathbb{C}$-linear combination of such. Since U is compact, $\|\alpha\|$ is bounded on U, so too is $\|\alpha\|^{s-1}$ for fixed $\operatorname{Re} s > 1$; thus

$$\int_{U\cap B^\times} \|\alpha\|^s\,\mathrm{d}\tau^\times(\alpha) = \int_{\mathcal{O}\cap B^\times} \|\alpha\|^{s-1}\,\mathrm{d}\tau(\alpha)$$

is bounded by a constant multiple of $\tau(U) < \infty$, and therefore $Z_B^{\Phi}(s)$ is bounded by comparison. ∎

This integral representation recovers the classical zeta function we studied earlier in the number field case (section 26.4).

29.7.16. Since there is a unique ideal in R of absolute norm q^e for each $e \geq 0$, we have

$$\zeta_F(s) = \sum_{e=0}^{\infty} \frac{1}{q^{es}} = (1 - q^{-s})^{-1}.$$

In like fashion, for $e \geq 0$, let $a_{\mathfrak{p}^e} = a_{\mathfrak{p}^e}(\mathcal{O})$ be the number of (necessarily principal) right ideals of $\mathcal{O}$ of reduced norm $\mathfrak{p}^e$ and thereby absolute norm q^{ne}. We then define

$$\zeta_B(s) := \sum_{I \subseteq \mathcal{O}} \frac{1}{\mathsf{N}(I)^s} = \sum_{e=0}^{\infty} \frac{a_{\mathfrak{p}^e}}{q^{nes}},$$

the sum as usual over all nonzero right ideals of $\mathcal{O}$.

Lemma 29.7.17. *The following statements hold.*

(a) *We have*

$$Z_B(s) = \tau^\times(O^\times)\zeta_B(s);$$

in particular, the domain of convergence of $Z_B(s)$ *is the same as* $\zeta_B(s)$*.*

(b) *If* B *is a quaternion algebra, then* $\zeta_B(s)$ *converges for* $\operatorname{Re} s > 1/2$*, and*

$$\zeta_B(s) = \begin{cases} \zeta_F(2s), & \text{if } B \text{ is a division algebra;} \\ \zeta_F(2s)\zeta_F(2s-1), & \text{if } B \simeq \mathrm{M}_2(F). \end{cases}$$

(c) *If* $B \simeq \mathrm{M}_n(F)$*, then* $\zeta_B(s)$ *converges for* $\operatorname{Re} s > 1 - 1/n$*, and*

$$\zeta_B(s) = \prod_{i=0}^{n-1} \zeta_F(ns - i).$$

Proof. First part (a). For $e \geq 0$, choose representatives for the $a_{\mathfrak{p}^e} = a_{\mathfrak{p}^e}(O)$ classes in $O/O^\times$ with reduced norm $\mathfrak{p}^e$. Every element $\alpha \in O \cap B^\times$ can be written as the product of one of the representatives and an element of $O^\times$; since $\|\alpha\| = |\mathrm{nrd}(\alpha)^n|$ we have

$$Z_B(s) = \int_{O\cap B^\times} \|\alpha\|^s \, \mathrm{d}\tau^\times(\alpha) = \tau^\times(O^\times) \sum_{e=0}^{\infty} \frac{a_{\mathfrak{p}^e}}{q^{nes}} = \tau^\times(O^\times)\zeta_B(s) \qquad (29.7.18)$$

as claimed. For (b) and (c), we gave a formula for $a_{\mathfrak{p}^e}$ for $B = \mathrm{M}_n(F)$ in Exercise 26.12 and for B a quaternion algebra in Corollary 26.4.7 (which extends to the function field case without change). Of course, these identities only hold within their respective domains of convergence, which for $\zeta_F(s) = (1 - q^{-s})^{-1}$ is $\operatorname{Re} s > 0$, thereby giving the rest. ∎

In the proof of the functional equation, we will need the following proposition, recalling duality (section 15.6).

Proposition 29.7.19. *For* Ψ *the characteristic function of* O*, we have*

$$\Psi^\vee(\alpha) = \begin{cases} \tau(O) = \mathsf{D}(B)^{-1/2}, & \text{if } \alpha \in O^\sharp; \\ 0, & \text{otherwise.} \end{cases}$$

Moreover,

$$Z_B^{\Psi^\vee}(s) = \mathsf{D}(B)^{s-1/2} Z_B^{\Psi}(s). \qquad (29.7.20)$$

Proof. For the first statement, by definition we have

$$\Psi^\vee(\alpha) = \int_O \psi(\alpha\beta)\,\mathrm{d}\tau(\beta).$$

If $\alpha \in O^\sharp$ then $\psi(\alpha\beta) = 1$ for all $\beta \in O$, and we obtain

$$\Psi^\vee(\alpha) = \tau(O) = \mathsf{D}(B)^{-1/2}\mu(O) = \mathsf{D}(B)^{-1/2}.$$

Otherwise, $\alpha \notin O^\sharp$, and by character theory $\Psi^\vee(\alpha) = 0$.

Since O is maximal, we have $O^\sharp = O\delta$ for some $\delta \in B^\times$ with $\|\delta\| = \mathsf{D}(B)$. Therefore

$$\begin{aligned} Z_B^{\Psi^\vee}(s) &= \mathsf{D}(B)^{-1/2} \int_{O^\sharp \cap B^\times} \|\alpha\|^s \, \mathrm{d}\tau^\times(\alpha) = \mathsf{D}(B)^{-1/2} \int_{O \cap B^\times} \|\alpha\delta\|^s \, \mathrm{d}\tau^\times(\alpha) \\ &= \mathsf{D}(B)^{s-1/2} \int_{O \cap B^\times} \|\alpha\|^s \, \mathrm{d}\tau^\times(\alpha) = \mathsf{D}(B)^{s-1/2} Z_B^{\Psi}(s), \end{aligned} \tag{29.7.21}$$

proving the second statement. □

We conclude this section with some hopefully illustrative computations of measure.

Lemma 29.7.22. *Let B be a quaternion algebra over F. Then*

$$\mu^\times(O^\times) = \begin{cases} 1 + 1/q, & \text{if } B \text{ is a division ring;} \\ 1 - 1/q^2, & \text{if } B \simeq \mathrm{M}_2(F) \end{cases} \tag{29.7.23}$$

Proof. If B is a division ring, then O is the valuation ring; let $J \subseteq O$ be the maximal ideal, so $O/J \simeq \mathbb{F}_{q^2}$ hence $\mu(J) = 1/q^2$, and then

$$\mu^\times(O^\times) = (1 - 1/q)^{-1} (\mu(O) - \mu(J)) = \frac{1 - 1/q^2}{1 - 1/q} = 1 + 1/q$$

Similarly, if $B \simeq \mathrm{M}_2(F)$ then $O \simeq \mathrm{M}_2(R)$, and from the exact sequence

$$1 \to 1 + \mathfrak{p}\,\mathrm{M}_2(R) \to \mathrm{GL}_2(R) \to \mathrm{GL}_2(k) \to 1$$

where $k = R/\mathfrak{p}$, we compute $\mu(1 + \mathfrak{p}\,\mathrm{M}_2(R)) = \mu(\mathfrak{p}\,\mathrm{M}_2(R)) = 1/q^4$ and

$$\begin{aligned} \mu^\times(\mathrm{GL}_2(R)) &= (1 - 1/q)^{-1} \mu(1 + \mathfrak{p}O) \# \mathrm{GL}_2(k) \\ &= \frac{(q^2 - 1)(q^2 - q)}{q^4(1 - 1/q)} = 1 - 1/q^2. \end{aligned}$$ □

29.7.24. Let B be a quaternion algebra. Then the reduced norm yields an exact sequence

$$1 \to O^1 \to O^\times \xrightarrow{\mathrm{nrd}} R^\times \to 1$$

so we have an induced compatible measure τ^1 on O^1. Thus

$$\begin{aligned} \tau^1(O^1) &= \frac{\tau^\times(O^\times)}{\tau^\times(R^\times)} = \frac{\mathsf{D}(B)^{-1/2}\mu^\times(O^\times)}{\mathsf{D}(R)^{-1/2}\mu^\times(R^\times)} \\ &= \frac{(d_F^4 \operatorname{discrd}(O)^2)^{-1/2}\mu^\times(O^\times)}{d_F^{-1/2}} \\ &= d_F^{-3/2}(1 - 1/q^2) \cdot \begin{cases} (q-1)^{-1}, & \text{if } B \text{ is a division algebra;} \\ 1, & \text{if } B \simeq \mathrm{M}_2(F). \end{cases} \end{aligned} \tag{29.7.25}$$

29.8 Idelic zeta functions

In this section, we now define (global) zeta functions in an idelic context. Let F be a global field and let B be a central simple algebra over F.

We first define the classical zeta function before generalizing to idelic zeta functions. We define

$$\zeta_B(s) := \prod_{v\nmid\infty} \zeta_{B_v}(s), \tag{29.8.1}$$

the product over all nonarchimedean places of the local zeta functions $\zeta_{B_v}(s)$ defined in 29.7.16. We studied these (global) zeta functions over number fields in Chapters 25 and 26.

Example 29.8.2. By Lemma 29.7.17(c) we have

$$\zeta_{\mathrm{M}_n(F)}(s) = \prod_{i=0}^{n-1} \zeta_F(ns-i).$$

With this in mind, we construct idelic zeta functions as products of local zeta functions, according to the product measures and characters, as follows.

29.8.3. For $v \in \operatorname{Pl} F$, we have defined measures μ_v on B_v in 29.5.1 for v archimedean and 29.7.1 for v nonarchimedean. We define the measure

$$\underline{\mu} := \prod_{v\in\operatorname{Pl} F} \mu_v$$

on $\underline{B}$ as the product measure (on the restricted direct product). In particular, for every eligible set $S \subset \operatorname{Pl} F$ of places and every $R_{(S)}$-order O in B, the measure $\underline{\mu}$ restricts to the (convergent) product measure on $\prod_{v\in S} B_v \times \prod_{v\notin S} O_v$—and this uniquely defines the $\underline{\mu}$ as a Haar measure on $\underline{B}$.

Definition 29.8.4. The **absolute discriminant** $\mathsf{D}(B)$ of B is defined as the product of the local absolute discriminants over all nonarchimedean places:

$$\mathsf{D}(B) := \prod_{v\nmid\infty} \mathsf{D}(B_v).$$

The absolute discriminant is well-defined because $\mathsf{D}(B_v) = 1$ for all but finitely many v.

Example 29.8.5. If F is number field, let $d_F \in \mathbb{Z}$ be the discriminant of F. Multiplying the factors 29.7.6 together, if B is a quaternion algebra we find that

$$\mathsf{D}(B) = \prod_{\mathfrak{p}\in\operatorname{Ram}(B)} \mathsf{N}(\mathfrak{p})^2 \cdot \begin{cases} |d_F|^4, & \text{if } F \text{ is a number field;} \\ 1, & \text{if } F \text{ is a function field;} \end{cases}$$

the product taken over all nonarchimedean ramified places.

29.8.6. In light of Proposition 29.7.7, we define the **Tamagawa measure** on $\underline{B}$ by

$$\underline{\tau} := \mathsf{D}(B)^{-1/2}\underline{\mu} = \prod_v \tau_v, \tag{29.8.7}$$

with $\tau_v = \mu_v$ for v archimedean and $\tau_v = \mathsf{D}(B_v)^{-1/2}\underline{\mu}_v$ defined in (29.7.8) for v nonarchimedean.

Remark 29.8.8. The Tamagawa measure has an intrinsic definition, independent of local normalizations [Weil82, Chapter II]. This intrinsic definition is, among other things, important for generalizations (which are quite broad, see Remark 29.11.9). It is then a nontrivial effort to unravel definitions to verify that the general notion of Tamagawa measure reduces to the one defined above. We content ourselves with the characterization 29.8.12 below.

29.8.9. If F is a number field, we define the product character

$$\underline{\psi} := \prod_v \psi_v$$

on $\underline{B}$. If F is a function field, we align our local characters as follows: let $\omega \in \Omega_{F/\mathbb{F}_q}$ be a nonzero meromorphic 1-form, and for a place $v \in \operatorname{Pl} F$ let

$$\psi_v(\alpha) = \exp(2\pi i \langle \operatorname{trd}(\alpha)\omega/\mathrm{d}t_v\rangle_{F_v})$$

where t_v is a uniformizer at t, and we again define $\underline{\psi} := \prod_v \psi_v$. In both cases, $\underline{\psi}(\underline{\alpha}\underline{\beta}) = \underline{\psi}(\underline{\beta}\underline{\alpha})$ for all $\underline{\alpha}, \underline{\beta} \in \underline{B}$, as this holds for ψ_v for all v.

In all cases, we define the function

$$\underline{\Phi} := \prod_v \Phi_v$$

where Φ_v is the characteristic function of a maximal order O_v for v nonarchimedean and Φ_v is the standard function (Definition 29.5.13) if v is archimedean.

Proposition 29.8.10. *The product character $\underline{\psi}$ has the following properties.*

(a) *The map*

$$\begin{aligned} \underline{B} &\xrightarrow{\sim} \underline{B}^\vee \\ \underline{\alpha} &\mapsto (\underline{\beta} \mapsto \underline{\psi}(\underline{\alpha}\underline{\beta})) \end{aligned} \tag{29.8.11}$$

is an isomorphism of topological groups.

(b) *The image of B under (29.8.11) is the group of unitary characters that are trivial on B.*

(c) *The product measure $\underline{\tau}$ is self-dual with respect to $\underline{\psi}$.*

Proof sketch. In the number field case, the statement follows from the local versions and the product formula; in the function field case, it follows similarly and reduces to the fact that the sum of the residues is zero. ∎

29.8.12. The Tamagawa measure $\underline{\tau}$ is then also characterized as the unique Haar measure on $\underline{B}$ that is self-dual with respect to the Fourier transform associated to the standard character.

29.8.13. Moving now from additive to multiplicative, we define the **normalized** Haar measure

$$\underline{\mu}^\times = \prod_v \mu_v^\times \tag{29.8.14}$$

on $\underline{B}^\times$, where $\mu_v^\times$ is defined in 29.5.5 for v archimedean and (29.7.3) for v nonarchimedean. Finally, we define the **Tamagawa measure** on $\underline{B}^\times$ by

$$\underline{\tau}^\times := \mathsf{D}(B)^{-1/2}\underline{\mu}^\times = \prod_v \tau_v^\times, \tag{29.8.15}$$

where $\tau_v^\times = \mu_v^\times$ for v archimedean and $\tau_v^\times = \mathsf{D}(B_v)^{-1/2}\mu_v^\times$ as in (29.7.10) for v nonarchimedean.

With measures finally (!) in hand, we can now idelically integrate functions against them.

Definition 29.8.16. A **Schwartz–Bruhat** function on $\underline{B}$ is a finite linear combination of functions $f : \underline{B} \to \mathbb{C}$ with $f = \prod_v f_v$, where each f_v is Schwartz–Bruhat and $f_v = \Psi_v$ is a standard function for all but finitely many v.

Theorem 29.8.17 (Poisson summation). *For a Schwartz–Bruhat function $\underline{\Phi}$, we have*

$$\sum_{\beta \in B} \underline{\Phi}(\beta) = \sum_{\beta \in B} \underline{\Phi}^\vee(\beta).$$

Proof. Symmetrize to obtain

$$(\Sigma\underline{\Phi})(\underline{\alpha}) := \sum_{\beta \in B} \underline{\Phi}(\underline{\alpha} + \beta). \tag{29.8.18}$$

Using $\underline{\psi}$ to identify B with the dual of $\underline{B}/B$ via Proposition 29.8.10 (and attending carefully to relevant analytic concerns), the symmetrized function is equal to its Fourier series

$$(\Sigma\underline{\Phi})(\underline{\alpha}) = \sum_{\beta \in B} a_\beta \underline{\psi}(-\beta\underline{\alpha}) \tag{29.8.19}$$

where

$$\begin{aligned} a_\beta &= \int_{\underline{B}/B} (\Sigma\underline{\Phi})(\underline{\alpha})\underline{\psi}(\beta\underline{\alpha})\,\mathrm{d}\underline{\tau}(\underline{\alpha}) \\ &= \int_{\underline{B}/B} \sum_{\gamma \in B} \underline{\Phi}(\underline{\alpha} + \gamma)\underline{\psi}(\beta(\underline{\alpha} + \gamma))\,\mathrm{d}\underline{\tau}(\underline{\alpha}) \\ &= \int_{\underline{B}} \underline{\Phi}(\underline{\alpha})\underline{\psi}(\beta\underline{\alpha})\,\mathrm{d}\underline{\tau}(\underline{\alpha}) = \underline{\Phi}^\vee(\beta) \end{aligned} \tag{29.8.20}$$

since $\underline{\psi}(\gamma) = 1$ for all $\gamma \in B$. Therefore

$$\sum_{\beta \in B} \underline{\Phi}(\beta) = (\Sigma\underline{\Phi})(0) = \sum_{\beta \in B} \underline{\Phi}^\vee(\beta).$$

□

Remark 29.8.21. When F is a function field, the alignment in $\underline{\psi}$ implies the Riemann–Roch theorem for the curve with function field F: see Exercise 29.13.

Definition 29.8.22. For a Schwartz–Bruhat function $\underline{\Phi}$, we define the **(idelic) zeta function**

$$Z_{\underline{B}}^{\underline{\Phi}}(s) := \int_{\underline{B}^\times} \underline{\Phi}(\underline{\alpha})\|\underline{\alpha}\|^s \, d\underline{\tau}^\times(\underline{\alpha});$$

when $\underline{\Phi} = \underline{\Psi}$ is a standard function, we write simply $Z_B(s)$.

We have

$$Z_{\underline{B}}^{\underline{\Phi}}(s) = \prod_{v \in \operatorname{Pl} F} Z_{B_v}^{\Phi_v}(s) \tag{29.8.23}$$

wherever the product is absolutely convergent; the local zeta functions $Z_{B_v}(s)$ are defined in (29.5.16) for archimedean v and (29.7.13) for nonarchimedean v.

We can make the comparison to the classical zeta function (29.8.1) explicit up to an idelic volume, as follows.

Lemma 29.8.24. *The following statements hold.*

(a) *If F is a number field, then*

$$Z_B(s) = \widehat{\tau}^\times(\widehat{O}^\times)\zeta_B(s) \prod_{v \mid \infty} Z_{B_v}(s);$$

if F is a function field, then

$$Z_B(s) = \underline{\tau}^\times(\underline{O}^\times)\zeta_B(s).$$

(b) *Suppose B is a quaternion algebra. If F is a number field with $\mathfrak{D} :=$ disc B, then*

$$\widehat{\tau}^\times(\widehat{O}^\times)^{-1} = |d_F|^2 \zeta_F(2)\varphi(\mathfrak{D}) \tag{29.8.25}$$

where $\varphi(\mathfrak{D}) = \prod_{\mathfrak{p} \mid \mathfrak{D}}(\operatorname{Nm}(\mathfrak{p}) - 1)$. If F is a function field, then

$$\underline{\tau}^\times(\underline{O}^\times)^{-1} = \zeta_F(2) \prod_{v \in \operatorname{Ram}(B)} (1 - 1/q_v).$$

Proof. For part (a), we use Lemma 29.7.17 for the relationship between zeta functions. For part (b), we combine Lemmas 29.5.18 and 29.7.17(b) and we use Lemma 29.7.22 for the local computation of measure and the equality

$$\prod_{\mathfrak{p} \mid \mathfrak{D}} (1 - \mathsf{N}(\mathfrak{p})^{-1})\sqrt{\mathsf{D}(B)} = |d_F|^2 \varphi(\mathfrak{D})$$

in the number field case. □

29.8.26. For the modulus $\| \;\|$, we find that

$$\|\underline{\alpha}\| = \prod_v \|\alpha_v\|_v.$$

In particular, recalling 27.6.11 we have the group

$$\underline{B}^{(1)} = \left\{\underline{\alpha} \in \underline{B}^\times : \|\underline{\alpha}\| = 1\right\},$$

and by the product formula we have $B^\times \leq \underline{B}^{(1)}$.

We now restrict the measure on $\underline{B}^{(1)}$ to $B^\times$. Let $n^2 = \dim_F B$. Then

$$\|\alpha\|_B = \|\mathrm{nrd}(\alpha)^n\|_F.$$

We have an exact sequence

$$1 \to \underline{B}^{(1)} \to \underline{B}^\times \to \|\underline{B}^\times\| \to 1; \tag{29.8.27}$$

we have $\|\underline{B}^\times\| = \mathbb{R}_{>0}$ if F is a number field and $\|\underline{B}^\times\| \leq q^{n\mathbb{Z}}$ if F is a function field with constant field $\mathbb{F}_q$. Noting this, we take the measure on $\|\underline{B}^\times\|$ defined by $n^{-1}\,\mathrm{d}t/t$ in both cases. By compatibility (see 29.3.13), we obtain a measure $\underline{\tau}^{(1)}$ on $\underline{B}^{(1)}$.

29.9 Convergence and residue

In this section, we establish convergence properties of idelic zeta functions, and compute their residue at $s = 1$. Throughout, F is a global field and B is a central simple algebra over F. Let $n^2 = \dim_F B$; we allow $n = 1$, so $B = F$.

Proposition 29.9.1. *Let $\underline{\Phi}$ be a Schwartz–Bruhat function on $\underline{B}$. Then the following statements hold.*

(a) $Z_{\underline{B}}^{\underline{\Phi}}(s)$ *is absolutely convergent for* $\mathrm{Re}(s) > 1$.

(b) $Z_{\underline{B}}^{\underline{\Phi}}(s)$ *has a simple pole at* $s = 1$ *with residue*

$$\mathrm{res}_{s=1}\, Z_{\underline{B}}^{\underline{\Phi}}(s) = \frac{\underline{\Phi}^\vee(0)\zeta_F^*(1)}{n}$$

where $\zeta_F^*(1) = \mathrm{res}_{s=1}\,\zeta_F(1)$.

Proof. In a nutshell, we prove this proposition by comparison to the classical zeta function $\zeta_{\mathrm{M}_n(F)}(s)$ of the matrix ring, for which we can compare explicitly with a Dirichlet series.

We may suppose without loss of generality that $\underline{\Phi} = \prod_v \underline{\Phi}_v$, since by definition $\underline{\Phi}$ is a linear combination of such. Let $S \subseteq \mathrm{Pl}(F)$ be the (finite) set of places v of F such that one of the following holds:

- v is archimedean (if F is a number field);
- $B_v \not\simeq \mathrm{M}_n(F_v)$; or
- $\underline{\Phi}_v \neq \underline{\Psi}_v$, i.e., $\underline{\Phi}_v$ is not the standard function.

Then

$$Z_B^\Phi(s) = \prod_{v\in S} Z_{B_v}^\Phi(s) \prod_{v\notin S} Z_{\mathrm{M}_n(F_v)}(s). \tag{29.9.2}$$

where the first product is finite. By Lemma 29.7.15 for $v \in S$ archimedean and comparison to Lemma 29.5.18 for v archimedean, the first product is absolutely convergent for $\operatorname{Re} s > 1$. For the second (infinite) product, by Lemma 29.7.17(c), we have

$$\zeta_{\mathrm{M}_n(F_v)}(s) = \prod_{i=0}^{n-1} \zeta_{F_v}(ns-i) = \prod_{i=0}^{n-1}(1-q^{i-ns})^{-1} \tag{29.9.3}$$

where q_v is the size of the residue field of F_v; note $\zeta_{\mathrm{M}_n(F_v)}(s)^{-1} = \prod_i (1-q^{i-ns})$ is holomorphic. Putting these together gives

$$\prod_{v\notin S} Z_{\mathrm{M}_n(F_v)}(s) = \widehat{\tau}^\times(\widehat{O}^\times)\zeta_{\mathrm{M}_n(F)}(s) \prod_{v\in S} Z_{\mathrm{M}_n(F_v)}(s)^{-1}.$$

and so we reduce to showing that $\zeta_{\mathrm{M}_n(F)}(s)$ is absolutely convergent for $\operatorname{Re}(s) > 1$. But multiplying (29.9.3) gives

$$\zeta_{\mathrm{M}_n(F)}(s) = \prod_{i=0}^{n-1} \zeta_F(ns-i)$$

which is absolutely convergent (by comparison to the harmonic series) whenever $n(\operatorname{Re} s) - (n-1) = n(\operatorname{Re} s) - n + 1 > 1$, i.e., when $\operatorname{Re} s > 1$. This proves part (a).

To prove part (b), we dig deeper. Staring at (29.9.3), we identify the terms $\zeta_{F_v}(ns-n+1)$ that contribute to a pole at $s = 1$; by absolute convergence, for $\operatorname{Re} s > 1$ we have

$$\frac{Z_B^\Phi(s)}{\zeta_F(ns-n+1)} = \prod_{v\mid\infty} Z_{B_v}^{\Phi_v}(s) \prod_{\substack{v\in S\\ v\nmid\infty}} \frac{Z_{B_v}^{\Phi_v}(s)}{\zeta_{F_v}(ns-n+1)} \prod_{v\notin S} \tau_v^\times(O_v^\times)\left(\prod_{i=0}^{n-2}\zeta_{F_v}(ns-i)\right). \tag{29.9.4}$$

The first two products are finite and defined at $s = 1$, and the third (infinite) product is now absolutely convergent for $\operatorname{Re} s > 1 - 1/n$, so in a neighorhood of $s = 1$. Accordingly, to compute the desired residue, we may compute a limit.

Let v be a nonarchimedean place of F and let q_v be the size of its residue field. For $\operatorname{Re} s > 1$, we have

$$\begin{aligned} Z_{B_v}^{\Phi_v}(s) &= \int_{B_v^\times} \Phi_v(\alpha_v)\|\alpha_v\|^s \,\mathrm{d}^\times\alpha_v \\ &= \left(\int_{B_v^\times} \Phi_v(\alpha_v)\|\alpha\|^{s-1}\,\mathrm{d}\alpha_v\right)(1-q_v^{-1})^{-1} \end{aligned} \tag{29.9.5}$$

coming from the normalization factors between additive and multiplicative Haar measure at the nonarchimedean places, where q_v is the size of the residue field of F_v.

Therefore

$$\begin{aligned}\lim_{s\searrow 1}\frac{Z_{B_v}^{\Phi_v}(s)}{\zeta_{F_v}(ns-n+1)} &= \lim_{s\searrow 1}\frac{1-q_v^{-ns+n-1}}{1-q_v^{-1}}\int_{B_v^\times}\Phi_v(\alpha_v)\|\alpha\|^{s-1}\,\mathrm{d}\alpha_v \\ &= \int_{B_v}\Phi_v(\alpha_v)\,\mathrm{d}\alpha_v = \Phi_v^\vee(0).\end{aligned} \tag{29.9.6}$$

In a similar way, if v is archimedean we have (without annoying normalization factors)

$$\lim_{s\searrow 1} Z_{B_v}^{\Phi_v}(s) = \int_{B_v}\Phi_v(\alpha_v)\,\mathrm{d}\alpha_v = \underline{\Phi}_v^\vee(0).$$

Applying (29.9.6), we conclude

$$\begin{aligned}\lim_{s\searrow 1}\frac{Z_{B}^{\Phi}(s)}{\zeta_{F}(ns-n+1)} &= \prod_{v\mid\infty}\lim_{s\searrow 1} Z_{B_v}^{\Phi_v}(s)\prod_{v\nmid\infty}\lim_{s\searrow 1}\frac{Z_{B_v}^{\Phi_v}(s)}{\zeta_{F_v}(ns-n+1)} \\ &= \prod_{v}\Phi_v^\vee(0) = \underline{\Phi}^\vee(0);\end{aligned} \tag{29.9.7}$$

the interchange of the product and the limit is justified by absolute convergence of the product in a neighborhood of $s=1$. Therefore

$$\operatorname{res}_{s=1} Z_B^{\Phi}(s) = \underline{\Phi}^\vee(0)\operatorname{res}_{s=1}\zeta_F(ns-n+1) = \frac{\underline{\Phi}^\vee(0)\zeta_F^*(1)}{n}$$

as claimed, finishing the proof of (b). □

29.10 Main theorem

We now establish the main analytic properties of the idelic zeta function, including meromorphic continuation and evaluation of residues. Our basic reference is Weil [Weil82, Section III.1]; see Remark 29.10.24 for historical comments and further references. The proof follows the same strategy as in section 29.2, with a key role played by Poisson summation and conceptual clarity brought by idelic methods.

Main Theorem 29.10.1. *Let F be a global field, let B be a central* division *algebra over F with $n^2 = \dim_F B$. Let $\underline{\Phi}$ be a Schwartz–Bruhat function on $\underline{B}$. Then the following statements hold.*

(a) *The function $Z_B^{\Phi}(s)$ has meromorphic continuation to $\mathbb{C}$. Moreover:*

(i) *If F is a number field, then $Z_B^{\Phi}(s)$ is holomorphic in $\mathbb{C}\smallsetminus\{0,1\}$ with simple poles at $s=0,1$ and residues*

$$\operatorname{res}_{s=1} Z_B^{\Phi}(s) = \frac{\underline{\Phi}^\vee(0)\zeta_F^*(1)}{n}, \qquad \operatorname{res}_{s=0} Z_B^{\Phi}(s) = -\frac{\underline{\Phi}(0)\zeta_F^*(1)}{n}.$$

(ii) *If F is a function field with field of constants $\mathbb{F}_q$, then $Z_{\underline{B}}^{\Phi}(s)$ is holomorphic in $\mathbb{C}\smallsetminus\{s : q^s = q^0, q^1\}$ with simple poles when $q^s = q^0, q^1$ and residues*

$$\operatorname{res}_{q^s=q^1} Z_{\underline{B}}^{\Phi}(s) = \frac{\underline{\Phi}^{\vee}(0)\zeta_F^*(1)}{n}, \quad \operatorname{res}_{q^s=q^0} Z_{\underline{B}}^{\Phi}(s) = -\frac{\underline{\Phi}(0)\zeta_F^*(1)}{n}.$$

(b) $Z_{\underline{B}}^{\Phi}(s)$ *satisfies the functional equation*

$$Z_{\underline{B}}^{\Phi}(1-s) = Z_{\underline{B}}^{\Phi^{\vee}}(s). \tag{29.10.2}$$

(c) *We have*

$$\tau^{(1)}(B^{\times}\backslash B^{(1)}) = \zeta_F^*(1).$$

The value of the residue $\zeta_F^*(1)$ is given by the analytic class number formula (Theorem 26.2.3). Taking $n = 1$ gives the following important special case (without requiring this as input).

Corollary 29.10.3. *Let F be a global field and let $\zeta_F(s) := \prod_{v\nmid\infty}(1 - q_v^{-s})^{-1}$, the product over all nonarchimedean places of F and q_v the cardinality of the residue field at v. Then the following statements hold.*

(a) *If F is a number field with absolute discriminant d_F, let $\xi_F(s)$ be the completed Dedekind zeta function as defined in (26.8.4). Then $\xi_F(s)$ is holomorphic in $\mathbb{C}\smallsetminus\{0,1\}$, with simple poles at $s = 0,1$, and satisfies the functional equation $\xi_F(1-s) = \xi_F(s)$.*

(b) *If F is a function field of genus g with field of constants $\mathbb{F}_q$, let $\xi_F(s) := q^{(g-1)s}\zeta_F(s)$. Then $\zeta_F(s)$ is holomorphic in $\mathbb{C}\smallsetminus\{s : q^s = q^0, q^1\}$, with simple poles when $q^s = q^0, q^1$, and satisfies the functional equation $\xi_F(1-s) = \xi_F(s)$.*

Proof. Taking $n = 1$ gives $B = F$; we take $\underline{\Phi} = \underline{\Psi} = \prod_v \underline{\Psi}_v$ to be the standard function.

In the number field case, we have $\xi(s) = |d_F|^{s/2} Z_F(s)$. By Lemma 29.8.24 and Lemma 29.5.18 for the archimedean contribution, we see that $\xi_F(s) = c_F|d_F|^{s/2}Z_F(s)$ where c_F is a constant depending on F. We accordingly conclude holomorphicity from Main Theorem 29.10.1(a). For the functional equation, by (29.7.20), we have $Z_{F_v}(s) = \mathsf{D}(F_v)^{s-1/2}Z_{F_v}(s)$. Taking the product and reading off Main Theorem 29.10.1(b) gives

$$\xi(1-s) = c_F|d_F|^{(1-s)/2}Z_F(1-s) = |d_F|^{(1-s)/2+s-1/2}Z_F(s) = c_F|d_F|^{s/2}Z_F = \xi(s).$$

The function field case in concluded in a similar but more direct manner. □

Main Theorem 29.10.1 extends to the case of a matrix algebra (over a division algebra), with some additional complications because of the existence of zerodivisors: see Theorem 29.10.20 for the case $B = \mathrm{M}_2(F)$. As we will see, the proof is close to uniform in the number field and function field cases; we have separated these two cases in the statement for clarity, so in particular the poles in the function field case occur at $s \in 2\pi i(\log q)\mathbb{Z}$ and $s \in 1 + 2\pi i(\log q)\mathbb{Z}$.

Remark 29.10.4. Weil [Weil82] chooses a different normalization, giving a functional equation relating s to $n-s$ and with residues $\underline{\Phi}^\vee(0)\zeta_F^*(1)$ at $s=0$ and $-\underline{\Phi}(0)\zeta_F^*(1)$ at $s=n$.

Proof of Main Theorem 29.10.1. We work with Tamagawa measure throughout, so to ease notation we abbreviate $\mathrm{d}^\times := \mathrm{d}\tau^\times$. To break up the integral, we define

$$\lambda(t) := \begin{cases} 1, & 0 < t < 1; \\ 1/2, & t = 1; \\ 0, & t > 1. \end{cases} \tag{29.10.5}$$

We break up $\underline{B}^\times$ by λ according to the norm: for $\underline{\alpha} \in \underline{B}^\times$, we define

$$\begin{aligned} f_+(\underline{\alpha}) &:= \lambda(\|\underline{\alpha}\|^{-1}), \\ f_-(\underline{\alpha}) &:= \lambda(\|\underline{\alpha}\|). \end{aligned} \tag{29.10.6}$$

Then $f_+(\underline{\alpha}^{-1}) = f_-(\underline{\alpha})$ and $f_+(\underline{\alpha}) + f_-(\underline{\alpha}) = 1$ for all $\underline{\alpha} \in \underline{B}^\times$, and so defining

$$Z_{\underline{B}}^{\Phi,\pm}(s) = \int_{\underline{B}^\times} f_\pm(\underline{\alpha})\|\underline{\alpha}\|^s \underline{\Phi}(\underline{\alpha})\,\mathrm{d}^\times\underline{\alpha}, \tag{29.10.7}$$

we obtain

$$Z_{\underline{B}}^{\Phi}(s) = Z_{\underline{B}}^{\Phi,+}(s) + Z_{\underline{B}}^{\Phi,-}(s). \tag{29.10.8}$$

We claim that the function $Z_{\underline{B}}^{\Phi,+}(s)$ is holomorphic. Indeed, by Proposition 29.9.1, $Z_{\underline{B}}^{\Phi}(s)$ converges absolutely for $\operatorname{Re} s > 1$; thus the same is true for $Z_{\underline{B}}^{\Phi,+}(s)$. But if $Z_{\underline{B}}^{\Phi,+}(s)$ converges absolutely at $s = s_0$, then it does so for all $\operatorname{Re}(s) \le \operatorname{Re}(s_0)$, so $Z_{\underline{B}}^{\Phi,+}(s)$ is holomorphic in $\mathbb{C}$.

For the remaining piece $Z_{\underline{B}}^{\Phi,-}(s)$, we have

$$Z_{\underline{B}}^{\Phi,-}(s) = \int_{B^\times\backslash\underline{B}^\times} f_-(\underline{\alpha})\|\underline{\alpha}\|^s \left(\sum_{\beta\in B^\times} \underline{\Phi}(\beta\underline{\alpha})\right)\mathrm{d}^\times\underline{\alpha}. \tag{29.10.9}$$

Poisson summation (Theorem 29.8.17) gives

$$\sum_{\beta\in B} \underline{\Phi}(\beta) = \sum_{\beta\in B} \underline{\Phi}^\vee(\beta).$$

For $\underline{\alpha} \in \underline{B}^\times$, let $\underline{\Phi}_{\underline{\alpha}}(\underline{\beta}) = \underline{\Phi}(\underline{\beta}\underline{\alpha})$ be the right translate. Then

$$\begin{aligned} \underline{\Phi}_{\underline{\alpha}}^\vee(\underline{\beta}) &= \int_{\underline{B}} \underline{\Phi}(\underline{\gamma}\underline{\alpha})\underline{\psi}(\underline{\beta}\underline{\gamma})\,\mathrm{d}\underline{\tau}(\underline{\gamma}) \\ &= \frac{1}{\|\underline{\alpha}\|}\int_{\underline{B}} \underline{\Phi}(\underline{\gamma}\underline{\alpha})\underline{\psi}(\underline{\beta}\underline{\gamma}\underline{\alpha}^{-1})\,\mathrm{d}\underline{\tau}(\underline{\gamma}) \quad = \frac{1}{\|\underline{\alpha}\|}\underline{\Phi}^\vee(\underline{\alpha}^{-1}\underline{\beta}) \end{aligned} \tag{29.10.10}$$

using that ψ is well-defined on conjugacy classes in the last equality. Plugging this into Theorem 29.8.17 gives

$$\sum_{\beta\in B} \underline{\Phi}(\beta\underline{\alpha}) = \frac{1}{\|\underline{\alpha}\|}\sum_{\beta\in B}\underline{\Phi}^\vee(\underline{\alpha}^{-1}\beta). \tag{29.10.11}$$

At this point, we use the hypothesis that B is a *division* algebra over F. With this assumption in hand,

$$\sum_{\beta\in B^\times} \underline{\Phi}(\beta\underline{\alpha}) = \sum_{\beta\in B}\underline{\Phi}(\beta\underline{\alpha}) - \underline{\Phi}(0).$$

From (29.10.11) applied to the sum in (29.10.9), we obtain

$$Z_{\underline{B}}^{\Phi,-}(s) = \int_{B^\times\backslash\underline{B}^\times} f_-(\underline{\alpha})\|\underline{\alpha}\|^{s-1}\left(\sum_{\beta\in B^\times}\underline{\Phi}^\vee(\underline{\alpha}^{-1}\beta) + \underline{\Phi}^\vee(0) - \underline{\Phi}(0)\|\underline{\alpha}\|\right)\mathrm{d}^\times\underline{\alpha} \tag{29.10.12}$$

valid for $\mathrm{Re}(s) > 1$.

Next, we make the substitution

$$s \leftarrow 1-s, \quad \underline{\alpha}\leftarrow\underline{\alpha}^{-1}$$

in the definition of $Z_{\underline{B}}^{\Phi^\vee,+}(s)$ (29.10.7). The Tamagawa measure $\mathrm{d}^\times\underline{\alpha}$ is invariant under inversion $\underline{\alpha}\leftarrow\underline{\alpha}^{-1}$, so

$$\begin{aligned} Z_{\underline{B}}^{\Phi^\vee,+}(1-s) &= \int_{\underline{B}^\times} f_+(\underline{\alpha}^{-1})\|\underline{\alpha}\|^{-(1-s)}\underline{\Phi}^\vee(\underline{\alpha}^{-1})\,\mathrm{d}^\times\underline{\alpha} \\ &= \int_{\underline{B}^\times} f_-(\underline{\alpha})\|\underline{\alpha}\|^{s-1}\underline{\Phi}^\vee(\underline{\alpha}^{-1})\,\mathrm{d}^\times\underline{\alpha}. \end{aligned} \tag{29.10.13}$$

We break up this integral according to its dependence on $B^\times$ as follows. We write

$$Z_{\underline{B}}^{\Phi^\vee,+}(1-s) = \int_{B^\times\backslash\underline{B}^\times} f_-(\underline{\alpha})\|\underline{\alpha}\|^{s-1}\left(\sum_{\beta\in B^\times}\underline{\Phi}^\vee(\underline{\alpha}^{-1}\beta)\right)\mathrm{d}^\times\underline{\alpha} \tag{29.10.14}$$

(replacing $\beta^{-1}\leftarrow\beta$ in the sum): in writing the integral this way, we integrate over any measurable set $\underline{B}^\times$ that maps injectively under the continuous quotient map $\underline{B}^\times\to B^\times\backslash\underline{B}^\times$. Combining (29.10.12) and (29.10.14), we obtain

$$Z_{\underline{B}}^{\Phi,-}(s) - Z_{\underline{B}}^{\Phi^\vee,+}(1-s) = \int_{B^\times\backslash\underline{B}^\times}\nu(\|\underline{\alpha}\|)\,\mathrm{d}^\times\underline{\alpha} \tag{29.10.15}$$

where

$$\nu(t) := \left(\underline{\Phi}^\vee(0)t^{s-1} - \underline{\Phi}(0)t^s\right)\lambda(t).$$

The function ν only depends on lengths in $\underline{B}^\times$. Recalling 29.8.26, in particular the exact sequence (29.8.27), we obtain

$$\int_{B^\times\backslash\underline{B}^\times}\nu(\|\underline{\alpha}\|)\,\mathrm{d}^\times\underline{\alpha} = \underline{\tau}^{(1)}(B^\times\backslash\underline{B}^{(1)})\left(\int_{\|\underline{B}^\times\|}\nu(t)\,\mathrm{d}^\times t\right) \tag{29.10.16}$$

where as in the previous paragraph,

$$\underline{\tau}^{(1)}(B^\times\backslash\underline{B}^{(1)}) = \int_{B^\times\backslash\underline{B}^{(1)}} \mathrm{d}\underline{\tau}^{(1)}\alpha$$

is the volume of any measurable set in $\underline{B}^{(1)}$ that maps injectively under the quotient map.

When F is a number field,

$$\begin{aligned}\int_{\|\underline{B}^\times\|} \nu(t)\,\mathrm{d}^\times t &= \frac{\underline{\Phi}^\vee(0)}{n}\int_0^1 t^{s-1}\,\frac{\mathrm{d}t}{t} - \frac{\underline{\Phi}(0)}{n}\int_0^1 t^s\,\frac{\mathrm{d}t}{t} \\ &= -\frac{1}{n}\left(\frac{\underline{\Phi}^\vee(0)}{1-s} + \frac{\underline{\Phi}(0)}{s}\right).\end{aligned} \tag{29.10.17}$$

When F is a function field with constant field $\mathbb{F}_q$,

$$\begin{aligned}\int_{\|\underline{B}^\times\|} \nu(t)\,\mathrm{d}t &= \frac{\underline{\Phi}^\vee(0)}{n}\left(\frac{1}{2} + \sum_{d=1}^{\infty} q^{-d(s-1)}\right) - \frac{\underline{\Phi}(0)}{n}\left(\frac{1}{2} + \sum_{d=1}^{\infty} q^{-ds}\right) \\ &= -\frac{1}{2n}\left(\underline{\Phi}^\vee(0)\frac{1+q^{s-1}}{1-q^{s-1}} + \underline{\Phi}(0)\frac{1+q^{-s}}{1-q^{-s}}\right).\end{aligned} \tag{29.10.18}$$

We now finish the proof, beginning with the case that F is a number field. Combining (29.10.8), (29.10.15), and (29.10.17), we obtain

$$\begin{aligned}Z_{\underline{B}}^{\Phi}(s) &= Z_{\underline{B}}^{\Phi,+}(s) + Z_{\underline{B}}^{\Phi^\vee,+}(1-s) \\ &\qquad - \frac{1}{n}\tau^{(1)}(B^\times\backslash B^{(1)})\left(\frac{\underline{\Phi}^\vee(0)}{1-s} + \frac{\underline{\Phi}(0)}{s}\right).\end{aligned} \tag{29.10.19}$$

The substitution $s \leftarrow 1-s$ interchanges $\underline{\Phi}$ and $\underline{\Phi}^\vee$, since $(\underline{\Phi}^\vee)^\vee(s) = \underline{\Phi}(s)$; from this we conclude the functional equation (b) and that $Z_{\underline{B}}^{\Phi}(s)$ has meromorphic continuation. Since $Z_{\underline{B}}^{\Phi,+}(s)$ and $Z_{\underline{B}}^{\Phi^\vee,+}(s)$ are entire, we conclude that $Z_{\underline{B}}^{\Phi}(s)$ is holomorphic in $\mathbb{C}$ away from $s = 0, 1$. Proposition 29.9.1(b) gives

$$\operatorname{res}_{s=1} Z_{\underline{B}}^{\Phi}(s) = \frac{\underline{\Phi}^\vee(0)\zeta_F^*(1)}{n}$$

which proves (a); on the other hand, (29.10.19) tells us

$$\operatorname{res}_{s=1} Z_{\underline{B}}^{\Phi}(s) = \frac{\tau^{(1)}(B^\times\backslash B^{(1)})\underline{\Phi}^\vee(0)}{n}$$

We conclude that $\tau^{(1)}(B^\times\backslash\underline{B}^{(1)}) = \zeta_F^*(1)$, which proves (c).

The case where F is a function field follows in a similar manner. □

The above extends to the case of a matrix algebra over a division algebra, with additional complications. To keep our eye on the prize, we treat just the case $B = \mathrm{M}_2(F)$.

Theorem 29.10.20. *Let F be a global field and let $B = \mathrm{M}_2(F)$. Let $\underline{\Phi}$ be a Schwartz–Bruhat function on $\underline{B}$. Then the conclusions of Main Theorem 29.10.1 hold with the exception that $Z_{\underline{B}}^{\Phi}(s)$ may also fail to be holomorphic at $s = 1/2$ (if F is a number field) or $q^s = q^{1/2}$ (if F is a function field).*

Proof. Let $B_1 := (a, b_1 \mid F)$ and $B_2 = (a, b_2 \mid F)$ be nonisomorphic division quaternion algebras with $\mathrm{Ram}(B_1) \cap \mathrm{Ram}(B_2) = \emptyset$, neither ramified at an archimedean place, and let $B_3 = (a, b_1 b_2 \mid F)$. Then by multiplicativity of the Hilbert symbol, we have $\mathrm{Ram}(B_3) = \mathrm{Ram}(B_1) \sqcup \mathrm{Ram}(B_2)$. For $i = 1, 2, 3$ and for each v, let $\Phi_{i,v} = \Phi_v$ if $v \notin \mathrm{Ram}(B_i)$ and $\Phi_{i,v} = \Psi_v$ the standard function if $v \in \mathrm{Ram}(B_i)$, and let $\underline{\Phi}_i = \prod_v \underline{\Phi}_{i,v}$. Then

$$Z_{\underline{B}}^{\Phi}(s) = \frac{Z_{\underline{B_1}}^{\Phi_1}(s) Z_{\underline{B_2}}^{\Phi_2}(s)}{Z_{\underline{B_3}}^{\Phi_3}(s)}. \tag{29.10.21}$$

The meromorphic continuation and functional equation for $Z_{\underline{B}}^{\Phi}$ then follow from the corresponding properties of $Z_{\underline{B_i}}^{\Phi_i}(s)$ and (29.10.21); we conclude also that $Z_{\underline{B}}^{\Phi}(s)$ has the claimed simple poles, and the values at the residues hold peeling off the factor $\zeta_F^*(1)/2$ because they hold at each v.

To conclude parts (a) and (b), we show that $Z_{\underline{B}}^{\Phi}(s)$ is holomorphic away from $s = 0, 1/2, 1$ (if F is a number field) and $q^s = q^{1/2}$ (if F is a function field). Decomposing the sum, we may suppose without loss of generality that $\Phi_v = \Psi_v$ is a standard function for all but finitely many places v. Let B' be a division quaternion algebra over F that is unramified at archimedean places and all places where $\Phi_v \neq \Psi_v$. Then

$$Z_{\underline{B}}^{\Phi}(s) = Z_{\underline{B'}}^{\Phi'}(s) \prod_{v \in \mathrm{Ram}(B')} \frac{Z_{B_v}^{\Psi_v}(s)}{Z_{B'_v}^{\Psi'_v}(s)}. \tag{29.10.22}$$

By Lemma 29.7.17, we have

$$\frac{Z_{B_v}^{\Psi_v}(s)}{Z_{B'_v}^{\Psi'_v}(s)} = \zeta_{F_v}(2s-1) = (1 - q_v^{1-2s})^{-1}$$

which is holomorphic away from $q_v^s = q_v^{1/2}$. When further F is a number field, we vary B' to conclude holomorphicity away from $s = 1/2$ (Exercise 29.10).

To keep things tidy, we conclude the proof of part (c) in Theorem 29.11.3. ◻

As an important special case, we have the following theorem, recalling the calculation of $\mathsf{D}(B)$ (Example 29.8.5).

Theorem 29.10.23. *Let B be a quaternion algebra over F and let $\mathsf{D}(B) \in \mathbb{Z}_{>0}$ be the absolute discriminant of B. Then the zeta function $Z_B(s)$ has the following properties.*

(a) *$Z_B(s)$ has meromorphic continuation to $\mathbb{C}$. More precisely:*

(i) *If F is a number field, then $Z_B(s)$ is holomorphic away from $s = 0, 1/2, 1$ with simple poles at $s = 0, 1$ and residues*

$$\operatorname{res}_{s=1} Z_B(s) = \frac{\zeta_F^*(1)}{2\sqrt{\mathsf{D}(B)}}, \quad \operatorname{res}_{s=0} Z_B(s) = -\frac{\zeta_F^*(1)}{2};$$

moreover, $Z_B(s)$ is holomorphic at $s = 1/2$ if and only if B is a division algebra.

(ii) *If F is a function field of a curve of genus g over $\mathbb{F}_q$, then $Z_B(s)$ is holomorphic away from $q^s = q^0, q^{1/2}, q^1$, with simple poles at $q^s = q^0, q^1$ and residues*

$$\operatorname{res}_{q^s=q^1} Z_B(s) = \frac{\zeta_F^*(1)}{2q^{2(2-2g)}\sqrt{\mathsf{D}(B)}}, \quad \operatorname{res}_{q^s=q^0} Z_B(s) = -\frac{\zeta_F^*(1)}{2};$$

moreover, $Z_B(s)$ is holomorphic where $q^s = q^{1/2}$ if and only if B is a division algebra.

(b) *$Z_B(s)$ satisfies a functional equation. More precisely:*

(i) *If F is a number field, then*

$$Z_B(1-s) = \mathsf{D}(B)^{1/2-s} Z_B(s).$$

(ii) *If F is a function field, then*

$$Z_B(1-s) = \left(q^{4(2-2g)}\mathsf{D}(B)\right)^{1/2-s} Z_B(s).$$

Proof. We apply Main Theorem 29.10.1 and Theorem 29.10.20 with $\underline{\Phi}$ as in 29.8.9 with Φ_v the characteristic function of a maximal order or the standard function according as v is nonarchimedean or archimedean, so $\underline{\Phi}(0) = 1$.

In the number field case, we have $\Phi_v^\vee = \Phi_v$ self-dual when v is archimedean, and by (29.7.20) we have

$$Z_{B_v}^{\Phi_v^\vee}(s) = \mathsf{D}(B_v)^{1/2-s} Z_B^{\Phi_v}(s)$$

and $\Phi_v^\vee(0) = \mathsf{D}(B_v)^{-1/2}$ when v is nonarchimedean. Multiplying these together and applying Theorem 29.10.1(c) gives

$$Z_{\underline{B}}^{\underline{\Phi}}(1-s) = Z_{\underline{B}}^{\underline{\Phi}^\vee}(s) = \mathsf{D}(B)^{1/2-s} Z_{\underline{B}}^{\underline{\Phi}}(s).$$

In the function field case, a similar argument holds but with the character modified by a global differential as in 29.8.9: for the relevant additional factor, see Exercise 29.13.

To conclude, we note that when $B \simeq \mathrm{M}_2(F)$ then by Lemma 29.8.24 the nonarchimedean part of $Z_B(s)$ is given by $\zeta_F(2s)\zeta_F(2s-1)$, and this has a double pole at $s = 1/2$ (or accordingly $q^{s-1/2} = 1$). □

Remark 29.10.24. In her 1929 Ph.D. thesis, Hey [Hey29] defined the zeta function of a division algebra over $\mathbb{Q}$, proving that it has an Euler product and functional equation—this was a *tour de force* in algebraic and analytic number theory, especially at the time!

For more on Hey's contribution and its role in the development of class field theory, see the perspective by Roquette [Roq2006, §9], including Zorn's observation that the functional equation yields an analytic proof of the classification of division algebras. Hey's thesis was never published (though the content appears in Deuring [Deu68, VII, §8]), and classical treatments of the zeta function for the most part gave way to the development of Chevalley's adeles and ideles.

Tate's thesis [Tate67] (from 1950, published in 1967) is usually given as the standard reference for the adelic recasting of zeta and L-functions of global fields as above: it gives the general definition of a zeta function associated to a local field, an integrable function, and a quasi-character [Tate67, §2]; see also the Bourbaki seminar by Weil [Weil66]. But already in 1946, Matchett (also a student of Emil Artin) wrote a Ph.D. thesis at Indiana University [Mat46] beginning the redevelopment of Hecke's theory of zeta and L-functions in terms of adeles and ideles. At the time, Iwasawa also contributed to the development of the theory; see his more recently published letter to Dieudonné [Iwa92].

These results were generalized to central simple algebras over local fields by Godement [God58a, God58b], Fujisaki [Fuj58], and Weil [Weil82, Chapter III] (and again Weil [Weil74, Chapter XI]) in the same style as Iwasawa and Tate; and then they were further generalized (allowing representations) by Godement–Jacquet [GJ72], providing motivation for the Langlands program.

As an example of the generalizations indicated by Remark 29.10.24, we conclude with a slightly more general statement.

Theorem 29.10.25. *Let F be a global field, let B be a central division algebra over F, let $\underline{\Phi}$ be a Schwartz–Bruhat function on $\underline{B}$, and let $\underline{\chi}\colon \underline{B}^\times \to \mathbb{C}$ be a unitary character such that $\underline{\chi}$ restricted to $\underline{F}^\times$ is trivial. Define*

$$L_B^{\Phi}(s,\underline{\chi}) := \int_{\underline{B}^\times} \underline{\Phi}(\underline{\alpha})\underline{\chi}(\underline{\alpha})\|\alpha\|^s \,\mathrm{d}\underline{\tau}^\times(\underline{\alpha}).$$

Then the function $L_B^{\Phi}(s,\underline{\chi})$ is absolutely convergent for $\operatorname{Re} s > 1$. *If $\underline{\chi}$ is nontrivial, then $L_B^{\Phi}(s,\underline{\chi})$ has* holomorphic *continuation to $\mathbb{C}$ and satisfies the functional equation*

$$L_B^{\Phi}(s,\underline{\chi}) = L_B^{\Phi^\vee}(s,\underline{\chi}^{-1}).$$

Proof. Proven in the same way as Main Theorem 29.10.1, just keeping track of the character $\underline{\chi}$. The (possible) residues at $s = 0, 1$ (or more generally in the function field case $q^s = q^0, q^1$) are multiplied by

$$\int_{\underline{B}^{(1)}/\underline{F}^\times} \underline{\chi}^{-1}(\underline{\alpha})\,\mathrm{d}\underline{\tau}^{(1)}(\underline{\alpha})$$

which is zero when $\underline{\chi}$ is nontrivial by character theory and therefore the L-function is fact holomorphic at these points. ∎

29.11 Tamagawa numbers

Building on Main Theorem 29.10.1, in this section we compute the measure of certain quotients with respect to the normalized idelic measure above. Let B be a quaternion algebra over the global field F.

Lemma 29.11.1. *Let Ω be the set of real ramified places in B, and let*

$$\underline{F}_{\Omega}^{(1)} := \{\underline{x} \in \underline{F}^{(1)} : x_v > 0 \textit{ for all } v \in \Omega\}.$$

Then the sequence

$$1 \to \underline{B}^1 \to \underline{B}^{(1)} \xrightarrow{\text{nrd}} \underline{F}_{\Omega}^{(1)} \to 1$$

is exact, giving a compatible measure on $\underline{B}^1$ denoted τ^1. Under this measure, we have

$$\underline{\tau}^{(1)}(B^\times \backslash \underline{B}^{(1)}) = \underline{\tau}^1(B^1 \backslash \underline{B}^1)\underline{\tau}^{(1)}(F^\times \backslash \underline{F}^{(1)}). \tag{29.11.2}$$

Proof. The surjectivity of the reduced norm is locally established, and its kernel is $\underline{B}^1$ by definition. The image of $B^\times$ under the reduced norm is $F^\times_{>_\Omega 0}$ by the Hasse–Schilling theorem of norms (Main Theorem 14.7.4), with kernel B^1. Moreover, the natural inclusion

$$F^\times_{>_\Omega 0} \backslash \underline{F}^{(1)}_{>_\Omega 0} \hookrightarrow F^\times \backslash \underline{F}^{(1)}$$

is also surjective by weak approximation. Putting these together, we obtain (29.11.2), which we might be tempted to summarize in the exact sequence of pointed sets

$$1 \to B^1 \backslash \underline{B}^1 \to B^\times \backslash \underline{B}^{(1)} \xrightarrow{\text{nrd}} F^\times \backslash \underline{F}^{(1)} \to 1$$

but we will resist the temptation. ∎

We now prove the final result in this chapter.

Theorem 29.11.3. *Let B be a quaternion algebra over a global field F. Then*

$$\underline{\tau}^{(1)}(B^\times \backslash \underline{B}^{(1)}) = \underline{\tau}^{(1)}(F^\times \backslash \underline{F}^{(1)}) = \zeta_F^*(1)$$

and

$$\underline{\tau}^1(B^1 \backslash \underline{B}^1) = 1.$$

Proof. The equality $\underline{\tau}^{(1)}(F^\times \backslash \underline{F}^{(1)}) = \zeta_F^*(1)$ is the statement of Main Theorem 29.10.1(c) applied to $B = F$. Then B is a division algebra, the equality $\underline{\tau}^{(1)}(B^\times \backslash \underline{B}^{(1)}) = \zeta_F^*(1)$ is again Main Theorem 29.10.1(c), and the equality $\underline{\tau}^1(B^1 \backslash \underline{B}^1) = 1$ then follows from (29.11.2).

To conclude, we make the appropriate modifications in the remaining case, and suppose that $B = \mathrm{M}_2(F)$. Then $B^1 = \mathrm{SL}_2(F)$ and similarly $\underline{B}^1 = \mathrm{SL}_2(\underline{F})$. By the exact sequence (29.11.2), we may show $\tau^1(\mathrm{SL}_2(F)\backslash \mathrm{SL}_2(\underline{F})) = \tau^1(\mathrm{SL}_2(\underline{F})/\mathrm{SL}_2(F)) = 1$.

We will do Fourier analysis on $\underline{F}^2$, extended from $\underline{F}$, with self-dual measure $\underline{\tau}$ and character ψ. Let $\underline{\Phi}$ be a Schwartz function on $\underline{F}^2$. The Fourier transform is

$$\underline{\Phi}^\vee(y) = \int_{\underline{F}^2} \underline{\Phi}(\underline{x})\psi(\underline{y}^{\mathsf{t}} \cdot \underline{x})\, \mathrm{d}\underline{x}$$

and Poisson summation reads

$$\sum_{x \in F^2} \underline{\Phi}(x) = \sum_{y \in F^2} \underline{\Phi}^\vee(y). \tag{29.11.4}$$

The group $\mathrm{SL}_2(\underline{F})$ acts on column vectors $\underline{F}^2 \smallsetminus \{0\}$ by left multiplication, with the stabilizer of $F^2 \smallsetminus \{0\}$ given by $\mathrm{SL}_2(F)$. Thus

$$\underline{\Phi}^\vee(0) = \int_{\underline{F}^2} \underline{\Phi}(\underline{x})\, \mathrm{d}\underline{\tau}(\underline{x}) = \int_{\mathrm{SL}_2(\underline{F})/\,\mathrm{SL}_2(F)} \left(\sum_{x \in F^2} \underline{\Phi}(\underline{\alpha} x) - \underline{\Phi}(0) \right) \mathrm{d}\underline{\tau}^1(\underline{\alpha}). \tag{29.11.5}$$

From (29.11.4), we derive

$$\sum_{x \in F^2} \underline{\Phi}(\underline{\alpha} x) = \frac{1}{\|\underline{\alpha}\|} \sum_{y \in F^2} \underline{\Phi}^\vee((\underline{\alpha}^{\mathsf{t}})^{-1} y) \tag{29.11.6}$$

and $\|\underline{\alpha}\|_{\mathrm{SL}_2(\underline{F})} = 1$. Plugging in (29.11.6) into (29.11.5), we have

$$\int_{\underline{F}^2} \underline{\Phi}(\underline{x})\, \mathrm{d}\underline{\tau}(\underline{x}) = \int_{\mathrm{SL}_2(\underline{F})/\,\mathrm{SL}_2(F)} \left(\sum_{y \in F^2} \underline{\Phi}^\vee((\underline{\alpha}^{\mathsf{t}})^{-1} y) - \underline{\Phi}(0) \right) \mathrm{d}\underline{\tau}^1(\underline{\alpha}). \tag{29.11.7}$$

Replacing $\underline{\Phi} \leftarrow \underline{\Phi}^\vee$ and $\alpha \leftarrow (\alpha^{\mathsf{t}})^{-1}$ (preserving the measure) in (29.11.7) gives

$$\underline{\Phi}(0) = \int_{\underline{F}^2} \underline{\Phi}^\vee(\underline{x})\, \mathrm{d}\underline{\tau}(\underline{x}) = \int_{\mathrm{SL}_2(\underline{F})/\,\mathrm{SL}_2(F)} \left(\sum_{y \in F^2} \underline{\Phi}(\underline{\alpha} y) - \underline{\Phi}^\vee(0) \right) \mathrm{d}\underline{\tau}^1(\underline{\alpha}). \tag{29.11.8}$$

Subtracting (29.11.8) from (29.11.5) gives

$$\begin{aligned} \underline{\Phi}^\vee(0) - \underline{\Phi}(0) &= \int_{\mathrm{SL}_2(\underline{F})/\,\mathrm{SL}_2(F)} \left(\underline{\Phi}^\vee(0) - \underline{\Phi}(0) \right) \mathrm{d}\underline{\tau}^1(\underline{\alpha}) \\ &= \underline{\tau}^1(\mathrm{SL}_2(\underline{F})/\,\mathrm{SL}_2(F))(\underline{\Phi}^\vee(0) - \underline{\Phi}(0)); \end{aligned}$$

choosing $\underline{\Phi}$ such that $\underline{\Phi}(0) \neq \underline{\Phi}^\vee(0)$, we obtain $\underline{\tau}^1(\mathrm{SL}_2(\underline{F})/\,\mathrm{SL}_2(F)) = 1$. □

Remark 29.11.9. We return to Remark 29.8.8. In general, there is a natural, intrinsic measure on the adelic points of a semisimple algebraic group G over a number field F, called the **Tamagawa measure**. With respect to the Tamagawa measure, the volume $\mathrm{vol}(\mathsf{G}(\underline{F})/\mathsf{G}(F))$ is finite, and the **Tamagawa number** of G (over F) is defined as

$$\tau(\mathsf{G}) := \mathrm{vol}(\mathsf{G}(\underline{F})/\mathsf{G}(F)).$$

For example, in the above we computed the volume for the group G associated to the group B^1 of quaternions of reduced norm 1.

In the late 1950s, Tamagawa defined the Tamagawa measure [Tam66]. Weil [Weil82] (based on notes from lectures at Princeton 1959–1960), computed that $\tau(G) = 1$ for G a *classical* semisimple and simply connected group; the conjecture that this holds in general was known as Weil's conjecture on Tamagawa numbers. This difficult conjecture was proven by the efforts of many people: see Scharlau [Scha2009, §2] for a history. In particular, the calculation of the Tamagawa number of the orthogonal group of a quadratic form recovers the **Smith–Minkowski–Siegel mass formula** that computes the mass of a genus of lattice.

Exercises

1. Let μ_p be the standard Haar measure on $\mathbb{Q}_p$.

   (a) Let $D(a, \delta) := \{x \in \mathbb{Q}_p : |x - a| < \delta\}$ be the open ball of radius $\delta \in \mathbb{R}_{>0}$ around $a \in \mathbb{Q}_p$. Compute the measure $\mu_p(D(a, \delta))$, and repeat with the closed ball of radius $\delta \in \mathbb{R}_{\geq 0}$.
   (b) Show that
   $$\int_{\mathbb{Z}_p} \log(|x|_p)\,\mathrm{d}\mu_p(x) = -\frac{\log p}{p-1}.$$

2. Let $F := \mathbb{F}_q(T)$; then F is the function field of $\mathbb{P}^1$, a curve of genus $g = 0$ and
   $$\zeta_F(s) = \frac{1}{(1 - q^{-s})(1 - q^{1-s})}.$$
   Let $B := \mathrm{M}_2(F)$. Verify directly that $\zeta_B(s)$ as defined in (29.8.1) satisfies $\zeta_B(1 - s) = q^{4-8s}\zeta_B(s)$, and then compare this with Theorem 29.10.23.

3. Let G be a locally compact, second countable topological group, and let $H \leq G$ be a subgroup.

   (a) Show that G/H is locally compact. *[Hint: Use Exercise 12.4.]*
   (b) If H is closed, show that G/H is Hausdorff (repeating Exercise 12.5).

   [We do not need G to be Hausdorff in this exercise.]

4. Let G be a Hausdorff, locally compact, second countable topological group.

   (a) Let μ, μ' be two Haar measures on G. Show that there exists $\kappa \in \mathbb{R}_{>0}$ such that $\mu' = \kappa\mu$.
   (b) Let μ be a left Haar measure on G. Show that $\mu(G) < \infty$ if and only if G is compact. *[Hint: Use Exercise 12.4.]*

▷ 5. Let G be a Hausdorff, locally compact, second countable topological group. Let μ be a left Haar measure on G. Show that the modular function $\Delta_G : G \to \mathbb{R}_{>0}$ is a homomorphism of groups.

▷ 6. Let A be a Hausdorff, locally compact, second countable topological ring, and let $a, b \in A^\times$. Show that
   $$\|ab\| = \|a\|\|b\|.$$

7. Consider the exact sequences

$$1 \to \{\pm 1\} \to \mathbb{R}^\times \xrightarrow{\phi} \mathbb{R}^\times_{>0} \to 1$$

where the map ϕ is given by either the quotient by ± 1 (equivalently, the absolute value) or by the map $x \mapsto x^2$. Equip $\mathbb{R}^\times$ and $\mathbb{R}^\times_{>0}$ with the standard Haar measure $\mathrm{d}x/|x|$. Compute the unique compatible measures on $\{\pm 1\}$ for the two choices of ϕ and show they differ by a factor 2.

8. Finish the proof of Lemma 29.5.18. *[Hint: It may help to use the Iwasawa decomposition (Proposition 33.4.2 and Lemma 36.2.7).]*

▹ 9. Prove Proposition 29.6.9, as follows. Let F be a nonarchimedean local field and let $\psi\colon F \to \mathbb{C}^1$ be a nontrivial unitary character of F, for example the standard unitary character. For each $x \in F$, let $\psi_x\colon F \to \mathbb{C}^1$ be defined by $\psi_x(y) = \psi(xy)$.

   (a) Show that ψ_x is again a unitary character of F and that ψ_x is trivial if and only if $x = 0$.

   (b) Show that
   $$\begin{aligned} \Psi\colon F &\to F^\vee \\ x &\mapsto \psi_x \end{aligned}$$
   defines a continuous, injective group homomorphism. *[Hint: recall that $F^\vee$ inherits a compact-open topology, so a basis of neighborhoods of the identity is given by $\{f \in F^\vee : f(K) \subseteq V\}$ for $K \subseteq F$ compact and $V \subseteq F^\vee$ an open neighborhood of 1. Given such K, V, show that there exists an open neighborhood $U \subseteq F$ of 0 such that $xK \subseteq \psi^{-1}V$ for all $x \in U$.]*

   (c) Show that $\Psi(F)$ is dense in $F^\vee$. *[Hint: if $\psi_x(y) = 0$ for all $x \in F$, then $y = 0$.]*

   (d) Prove that Ψ^{-1} is continuous (on its image). *[Hint: work in the other direction in (b).]*

   (e) Show that $\Psi(F)$ is complete, hence closed, subgroup of $F^\vee$. Conclude that Ψ is an isomorphism of topological groups.

▹ 10. Complete the proof of Theorem 29.10.20 for F a number field, as follows. By choice of B', we showed that $Z_B^\Phi(s)$ is holomorphic except for $s = 0, 1$ and possibly when $q_v^s = q_v^{1/2}$ for $v \in \mathrm{Ram}(B')$. Show (by varying the choice of B') that $Z_B^\Phi(s)$ is holomorphic away from $s = 0, 1/2, 1$.

▹ 11. Let B be a quaternion algebra over a global field F. Show that there exists a compact $\underline{E} \subseteq \underline{B}$ such that the map $\underline{B} \to B\backslash\underline{B}$ is *not* injective on $\underline{E}$. *[Hint: accept the results of section 29.3 and take $\underline{E}$ with measure $\underline{\mu}(\underline{E}) > \underline{\mu}(B\backslash\underline{B})$ and integrate.]*

12. Following the proof of Theorem 29.11.3, show that if $B = \mathrm{M}_n(F)$ that

$$\underline{\tau}^{(1)}(B^\times\backslash\underline{B}^{(1)}) = \underline{\tau}^{(1)}(F^\times\backslash\underline{F}^{(1)}) = \zeta_F^*(1)$$

and

$$\underline{\tau}^1(B^1\backslash\underline{B}^1) = 1.$$

13. This exercise assumes some background in algebraic curves, see e.g. Silverman [Sil2009, Chapter II]—for a wider survey, see Rosen [Ros2002]. Let F be the function field of a curve X over $\mathbb{F}_q$. The **divisor group** $\operatorname{Div} X$ is the free abelian group on the set of places v; it has a degree map $\deg\colon \operatorname{Div} X \to \mathbb{Z}$ with $\deg v = [k_v : \mathbb{F}_q]$, where k_v is the residue field at v An element $f \in F^\times$ has a divisor $\operatorname{div} f = \sum_v \operatorname{ord}_v(f)v$. A nonzero meromorphic differential ω on X has also a divisor $K := \sum_v a_v v$ where $a_v = \operatorname{ord}_v(\omega/\mathrm{d}t_v)$ where t_v is a uniformizer at v; we call K a **canonical divisor**. Given $D = \sum_v d_v v \in \operatorname{Div} X$, let
$$\underline{L}(D) := \prod_v \mathfrak{p}_v^{-d_v} \subseteq \underline{F},$$
and $L(D) := \underline{L}(D) \cap F = \{f \in K : \operatorname{ord}_v(f) \geq -d_v\}$ and finally $\ell(D) := \dim_{\mathbb{F}_q} L(D)$. Define the **genus** of X or of F by $g := \ell(K) \in \mathbb{Z}_{\geq 0}$.
    (a) Show that the characteristic function $\underline{\Phi}$ of $\underline{L}(D)$ is $q^{\deg D - \deg K/2}$ times the characteristic function of $\underline{L}(K - D)$.
    (b) Apply Poisson summation to $\underline{\Phi}$ to obtain
$$\sum_{f \in L(D)} 1 = q^{\deg D - \deg K/2} \sum_{f \in L(K-D)} 1$$
$$\ell(D) = \deg D - \frac{1}{2}\deg K + \ell(K - D).$$
    (c) Plug in $D = 0$ to (b) to get $L(0) = \mathbb{F}_q$ the constant functions and conclude $\deg K = 2g - 2$.
    (d) Conclude the **Riemann–Roch theorem**
$$\ell(D) - \ell(K - D) = \deg D + 1 - g. \tag{29.11.10}$$
14. Prove Theorem 29.10.25.

This chapter is licensed under the terms of the Creative Commons Attribution-NonCommercial 4.0 International License (http://creativecommons.org/licenses/by-nc/4.0/), which permits any noncommercial use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license and indicate if changes were made.

The images or other third party material in this chapter are included in the chapter's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the chapter's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.

Chapter 30

Optimal embeddings

To conclude our analytic part, we apply idelic methods to understand embeddings of quadratic orders into quaternion orders.

30.1 ▹ Representation numbers

A subject of classical (and continuing) interest is the number of representations of an integer by an integral quadratic form, in particular as a sum of squares. Because of the subject of this text, we consider quadratic forms in three and four variables where quaternions provide insight.

Lagrange famously proved that every positive integer is the sum of four squares. We proved Lagrange's theorem (Theorem 11.4.3) by viewing the sum of four squares as the reduced norm on the Lipschitz order, and concluded the argument by comparison to the Hurwitz order, which is Euclidean. In Theorem 14.3.8, we proved Legendre's three-square theorem: every integer $n \geq 0$ not of the form $4^a(8b+7)$ is the sum of three squares. The proof is harder for three squares than for four (see Remark 11.4.4): our proof used the Hasse–Minkowski theorem (the local-global principle for quadratic forms over $\mathbb{Q}$) and again the fact that the Hurwitz order has class number 1. (We gave a variant in Exercise 14.5, where we used the local-global principle for embeddings.)

In each of these cases, we may also ask for a count of the *number* of such representations. For $k \in \mathbb{Z}_{\geq 1}$ and $n \in \mathbb{Z}_{\geq 0}$, let

$$r_k(n) := \#\{(x_1, x_2, \dots, x_k) \in \mathbb{Z}^k : x_1^2 + x_2^2 + \dots + x_k^2 = n\} \tag{30.1.1}$$

be the number of ways of writing n as the sum of k squares; equivalently, this is the number of lattice points on the sphere of radius $\sqrt{n}$ in $\mathbb{R}^k$. The number $r_4(n)$ is computed in terms of the factorization of n in the Hurwitz (or Lipschitz) orders, and has a simple answer: we saw in Exercise 11.3 that for an odd prime p,

$$r_4(p) = 8(p+1)$$

and we upgraded this in Exercise 26.7 to a general formula for $r_4(n)$ in terms of the sum of (odd) divisors of n.

© The Author(s) 2021
J. Voight, *Quaternion Algebras*, Graduate Texts in Mathematics 288,
https://doi.org/10.1007/978-3-030-56694-4_30

We may similarly ask for a formula for r_3, but it is more difficult both to state and to prove. Define

$$r_3^{\text{prim}}(n) = \#\{(x,y,z) \in \mathbb{Z}^3 : x^2 + y^2 + z^2 = n \text{ and } \gcd(x,y,z) = 1\} \tag{30.1.2}$$

as the number of **primitive** representations of n as the sum of three squares. Then $r_3^{\text{prim}}(n) = r_3(n)$ if n is squarefree, and more generally

$$r_3(n) = \sum_{d^2 \mid n} r_3^{\text{prim}}(n/d^2).$$

Let $h(d) = \#\operatorname{Pic} S_d$ be the class number of the quadratic order of discriminant $d < 0$; equivalently, $h(d)$ is the number of reduced primitive integral positive definite binary quadratic forms of discriminant d.

Gauss [Gau86, Section 291] showed that $r_3^{\text{prim}}(n)$ is a constant multiple of $h(-4n)$ as follows.

Theorem 30.1.3 (Gauss). *We have $r_3(1) = 6$, $r_3(3) = 8$, and for $n \in \mathbb{Z}_{\geq 0}$:*

$$r_3^{prim}(n) = \begin{cases} 0, & \text{if } n \equiv 0,4,7 \pmod 8; \\ 12h(-4n), & \text{if } n \equiv 1,2,5,6 \pmod 8 \text{ and } n \neq 1; \\ 8h(-4n) = 24h(-n), & \text{if } n \equiv 3 \pmod 8 \text{ and } n \neq 3. \end{cases}$$

(One can uniformly include the cases $n = 1, 3$ by accounting for the extra roots of unity in $\mathbb{Q}(\sqrt{-n})$.)

Theorem 30.1.3 is a special case of Theorem 30.4.7—see Exercise 30.4—but for historical and motivational reasons, we also give in the next section an essentially self-contained proof for the case $n \equiv 1, 2 \pmod 4$, following Venkov [Ven22, Ven29].

The main obstacle to generalizing Gauss's theorem (Theorem 30.1.3) is that quaternion orders need not have class number 1: a generalization with this hypothesis "following the general plan described by Venkov" is given by Shemanske [Shem86]. Another annoyance is the growing technicality of the local computations giving the explicit constants involved. Both of these issues are in some sense resolved by employing idelic methods (hence the placement of this chapter in this text) and even the proof of Gauss's theorem itself is simplified by these methods (in the next section). The result is Theorem 30.4.7: representations are spread across the genus of an order, with the constants given by local factors (computed in this chapter for maximal orders and then Eichler orders).

30.1.4. For indefinite quaternion orders, strong approximation applies, and we are almost always able to prove that the contribution to each order in the genus is equal, with one quite subtle issue known as *selectivity*: in certain rare circumstances, a quadratic order embeds in precisely *half* of the orders in a genus. We pursue selectivity in the next chapter (Chapter 31): technical and rather extraordinary, it is a subject that demands care.

Happily, a locally norm-maximal order (such as an Eichler order) over $\mathbb{Z}$ is not selective!

An important application of this theory is a refinement of the mass formula to a class number formula. Recall the Eichler mass formula (Theorem 25.3.18): if B is a definite quaternion algebra over $\mathbb{Q}$ of discriminant D and $O \subset B$ is an Eichler order of level M, then

$$\sum_{[J]\in \operatorname{Cls} O} \frac{1}{w_J} = \frac{\varphi(D)\psi(M)}{12}$$

where $w_J = \#O_{\mathsf{L}}(J)/\{\pm 1\}$. We can account for the necessary correction:

$$\#\operatorname{Cls} O = \sum_{[J]\in \operatorname{Cls} O} 1 = \frac{\varphi(D)\psi(M)}{12} + \sum_{\substack{[J]\in \operatorname{Cls} O \\ w_J > 1}} \left(1 - \frac{1}{w_J}\right).$$

The latter "error term" is accounted for by (finite) cyclic subgroups spread across representative left orders $O_{\mathsf{L}}(J)$—and this is precisely the contribution computed idelically above!

Theorem 30.1.5 (Eichler class number formula). *Let B be a definite quaternion algebra over $\mathbb{Q}$ of discriminant D, and let $O \subset B$ be an Eichler order of level M. Let $N = DM = \operatorname{discrd} O$. Then*

$$\#\operatorname{Cls} O = \frac{\varphi(D)\psi(M)}{12} + \frac{\epsilon_2}{4} + \frac{\epsilon_3}{3}$$

where

$$\epsilon_2 = \begin{cases} \displaystyle\prod_{p\mid D}\left(1 - \left(\frac{-4}{p}\right)\right)\prod_{p\mid M}\left(1 + \left(\frac{-4}{p}\right)\right), & \text{if } 4 \nmid N; \\ 0, & \text{if } 4 \mid N; \end{cases}$$

and

$$\epsilon_3 = \begin{cases} \displaystyle\prod_{p\mid D}\left(1 - \left(\frac{-3}{p}\right)\right)\prod_{p\mid M}\left(1 + \left(\frac{-3}{p}\right)\right), & \text{if } 9 \nmid N; \\ 0, & \text{if } 9 \mid N. \end{cases}$$

30.2 Sums of three squares

In this section, we prove the theorem of Gauss showing that the number of representations of an integer as a sum of three squares is a class number. For further reading, see also Grosswald [Gro85, Chapter 4] and the references therein.

A few parts are easy to establish. The count $r_3(1) = 6$ is immediate. If $4 \mid n$, then $r_3^{\text{prim}}(n) = 0$ since $x^2 + y^2 + z^2 \equiv 0 \pmod 4$ implies $x \equiv y \equiv z \equiv 0 \pmod 2$. Similarly, if $n \equiv 7 \pmod 8$ then $r_3^{\text{prim}}(n) = r_3(n) = 0$ by the three-square theorem.

The two remaining cases lie deeper. A proof using quaternions is due to Venkov [Ven22, Ven29]; alternate accounts are given by Hanlon [Hanlon81, Chapter 2] and Rehm [Reh76]. To accomplish the task of giving an argument that is as self-contained as possible and still previews the ideas and structure contained in this chapter, we give a proof in the case $n \equiv 1, 2 \pmod 4$.

Proof of Theorem 30.1.3 for $n \equiv 1, 2 \pmod 4$. Suppose $n \equiv 1, 2 \pmod 4$. Let $S := \mathbb{Z}[\sqrt{-n}] \subset K := \mathbb{Q}(\sqrt{-n})$. Then S is maximal and ramified at 2, i.e., $S \otimes \mathbb{Z}_2$ is the ring of integers of the field $K \otimes \mathbb{Q}_2 = \mathbb{Q}_2(\sqrt{-n})$.

Let $B = (-1, -1 \mid \mathbb{Q})$ be the rational Hamiltonians and $O \subset B$ the Hurwitz order. We consider the set

$$W = \{\alpha = xi + yj + zk \in O : \alpha \text{is primitive and} \operatorname{nrd}(\alpha) = n\}; \tag{30.2.1}$$

then $\#W = r_3^{\text{prim}}(n)$. By the three-square theorem, $W \neq \emptyset$, so let $\alpha \in W$. We embed

$$\begin{aligned} K &\hookrightarrow B \\ \sqrt{-n} &\mapsto \alpha \end{aligned}$$

and for convenience identify K with its image. By the Skolem–Noether theorem (Corollary 7.1.5), every other element $\alpha' \in W$ is the form $\alpha' = \beta^{-1}\alpha\beta$ with $\beta \in B^\times$ (but not necessarily conversely!). Let

$$E := \{\beta \in B^\times : \beta^{-1}\alpha\beta \in W\}. \tag{30.2.2}$$

The set E has a right action of the normalizer $N_{B^\times}(O)$ (checking that primitivity is preserved).

We relate E to the group of fractional ideals $\operatorname{Idl} S$ as follows. Let $\mathfrak{b} \subseteq K$ be a fractional S-ideal. Since O is (right) Euclidean, $\mathfrak{b}O = \beta O$ for some $\beta \in B^\times$ that is well-defined up to right multiplication by $O^\times$. The heart of the proof is the following claim: the map

$$\begin{aligned} \operatorname{Idl} S &\to E/N_{B^\times}(O) \\ \mathfrak{b} &\mapsto \beta N_{B^\times}(O) \end{aligned} \tag{30.2.3}$$

is a well-defined, surjective map of sets. The most efficient (and clear) proof of this claim is idelic, and we prove it in two steps.

First, the map (30.2.3) is well-defined: that is to say, $\beta \in E$. Write $\mathfrak{b}\widehat{S} = \widehat{b}\widehat{S}$, so that $\beta\widehat{O} = \widehat{b}\widehat{O}$ and $\beta = \widehat{b}\widehat{\mu}$ with $\widehat{\mu} \in \widehat{O}^\times$. Then $\widehat{b}$ commutes with α (in $\widehat{K} \subseteq \widehat{B}$), so

$$\alpha' = \beta^{-1}\alpha\beta = \widehat{\mu}^{-1}\widehat{b}^{-1}\alpha\widehat{b}\widehat{\mu} = \widehat{\mu}^{-1}\alpha\widehat{\mu} \in \widehat{\mu}^{-1}O\widehat{\mu} \subseteq \widehat{\mu}^{-1}\widehat{O}\widehat{\mu} \cap B = \widehat{O} \cap B = O.$$

To show that the map is surjective, we need to establish one other important point comparing the global and the idelic: we claim that there exists $\widehat{\nu} \in N_{\widehat{B}^\times}(\widehat{O})$ such that

$$\widehat{\nu}^{-1}\alpha\widehat{\nu} = \beta^{-1}\alpha\beta. \tag{30.2.4}$$

The existence of $\widehat{\nu}$ may be established locally. We prove this in Proposition 30.5.3: for $p \neq 2$, it amounts to showing that two elements of $\mathrm{M}_2(\mathbb{Z}_p)$ with square $-n$ are conjugate under $\mathrm{GL}_2(\mathbb{Z}_p)$, and for $p = 2$ it follows from the description of the valuation ring: it is here where we use the fact that S is maximal at 2, but this is only a technical issue. (The reader may accept (30.2.4) and proceed, or pause here and work this out, in Exercise 30.3.)

Given (30.2.4), we see that $\beta\widehat{\nu}^{-1}$ centralizes α in $B^\times$ and so belongs to $\widehat{K}^\times$, whence

$$\beta \in \widehat{K}^\times N_{\widehat{B}^\times}(\widehat{O}).$$

Finally, since 2 is ramified in K—using $n = 1, 2 \pmod 4$ again—and the nontrivial class in $N_{\widehat{B}^\times}(\widehat{O})/\widehat{O}^\times \simeq \mathbb{Z}/2\mathbb{Z}$ is represented by an element of reduced norm 2, we have

$$\widehat{K}^\times N_{\widehat{B}^\times}(\widehat{O}) = \widehat{K}^\times \widehat{O}^\times,$$

and therefore $\beta = \widehat{\mathfrak{b}}\widehat{\mu}$ with $\widehat{\mathfrak{b}} \in \widehat{K}^\times$ and $\widehat{\mu} \in \widehat{O}^\times$. Therefore $\widehat{\mathfrak{b}}\widehat{O} = \beta\widehat{O}$ and so $\beta O = \mathfrak{b}O$ where $\mathfrak{b} = \widehat{\mathfrak{b}}\widehat{S}$.

From the claim, we conclude the theorem. We have $\beta^{-1}\alpha\beta = \alpha$ if and only if β centralizes α if and only if $\beta \in K^\times$, so the desired elements $\beta^{-1}\alpha\beta$ up to $N_{B^\times}(O)$ are uniquely determined by the class $[\mathfrak{b}] \in \operatorname{Pic} S = \operatorname{Idl} S/\operatorname{PIdl} S$. Finally, when $n > 1$ we have $S^\times = \{\pm 1\}$ so $\#O^\times/S^\times = 12$, and the result follows. ◻

Importantly, Venkov's proof of Gauss's theorem given above is explicit and constructive, given at least one representation as a sum of three squares.

The early observation made in the proof above is that the sum of three squares is the restriction of the reduced norm to the trace zero elements of the Hurwitz order. One then seeks a similar statement for quadratic forms $Q = \operatorname{nrd}|_{O^0}$ obtained more generally. (This is *almost* the same thing as the ternary quadratic form associated to O itself via the Clifford algebra construction in Chapter 22; the difference is that for the latter we take the *dual* of the order and scale, as in (22.1.3).)

Just as in the proof above, for a representation $Q(x, y, z) = n$ corresponding to $\alpha \in O$ with $\alpha^2 + n = 0$, we obtain an embedding $S = \mathbb{Z}[\sqrt{-n}] \hookrightarrow O$ of a quadratic order into the quaternion order; conversely, to such an embedding we find a representation. It is more convenient to work with embeddings, as they possess more structure. Viewed in this way, we may equivalently restrict the reduced norm from O to the order S itself, and then we are asking for the representation of a *binary* quadratic form by a quaternary quadratic form.

30.3 Optimal embeddings

We now begin in earnest. We start by considering quadratic embeddings into quaternions, both rationally and integrally. Let R be a Dedekind domain with $F = \operatorname{Frac} R$, and let B be a quaternion algebra over F.

Let K be a separable quadratic F-algebra: then either $K \supseteq F$ is a separable quadratic field extension or $K \simeq F \times F$. Suppose that $K \hookrightarrow B$.

30.3.1. The set of embeddings of K in B is identified with the set $K^\times\backslash B^\times$, by 7.7.12: if $\phi\colon K \hookrightarrow B$ is another embedding, then by the Skolem–Noether theorem there exists $\beta \in B^\times$ such that $\phi(\alpha) = \beta^{-1}\alpha\beta$ for all $\alpha \in K$ with β well-defined up to left multiplication by $K^\times$, the centralizer of K under conjugation by $B^\times$.

We now turn to the integral theory. Let $O \subseteq B$ be a quaternion R-order, and let $S \subseteq K$ be a quadratic R-order; we will be interested in embeddings $\phi\colon S \hookrightarrow O$. Such an embedding gives an embedding $\phi\colon K \hookrightarrow B$ by extending scalars. We keep the embeddings for various suborders organized as follows.

Definition 30.3.2. An R-algebra embedding $\phi\colon S \hookrightarrow O$ is **optimal** if

$$\phi(K) \cap O = \phi(S).$$

Let

$$\mathrm{Emb}_R(S, O) := \{\text{Optimal embeddings } S \hookrightarrow O\}. \tag{30.3.3}$$

When no confusion can result, we drop the subscript R and write simply $\mathrm{Emb}(S, O)$.

30.3.4. If an embedding $\phi\colon S \hookrightarrow O$ is not optimal, then it is optimal for a larger order $S' \supseteq S$. Accordingly, there is a natural decomposition

$$\{\text{Embeddings } S \hookrightarrow O\} = \bigsqcup_{S' \supseteq S} \mathrm{Emb}(S', O). \tag{30.3.5}$$

Lemma 30.3.6. *An R-algebra embedding $\phi\colon S \hookrightarrow O$ is optimal if and only if the induced $R_{\mathfrak{p}}$-algebra embeddings $S_{\mathfrak{p}} \hookrightarrow O_{\mathfrak{p}}$ are optimal for all primes $\mathfrak{p} \subseteq R$.*

Proof. Immediate from the local-global dictionary for lattices (Theorem 9.4.9). □

Lemma 30.3.6 says that for an embedding ϕ, the property of being optimal is a local property.

We define

$$\begin{aligned} E = E_{S,O} &:= \{\beta \in B^\times : \beta^{-1} K \beta \cap O = \beta^{-1} S \beta\} \\ &= \{\beta \in B^\times : K \cap \beta O \beta^{-1} = S\}. \end{aligned} \tag{30.3.7}$$

(The set E also depends on the fixed embedding $K \hookrightarrow B$, but this ‘reference’ embedding will remain fixed throughout.) In equation 30.3.7, we see two different ways to think about embeddings: we either move S and see how it fits into O, or we fix K and move O.

Lemma 30.3.8. *The map*

$$\begin{aligned} K^\times \backslash E &\xrightarrow{\sim} \mathrm{Emb}(S, O) \\ \beta &\mapsto \phi_\beta \end{aligned} \tag{30.3.9}$$

where $\phi_\beta(\alpha) := \beta^{-1}\alpha\beta$ for $\alpha \in S$ is a bijection.

Proof. Immediate from 30.3.1. □

We further organize our optimal embeddings up to conjugation as follows.

30.3.10. Let $O^1 \leq \Gamma \leq N_{B^\times}(O)$. Then the image of Γ in $N_{B^\times}(O)/F^\times$ has finite index. For example, we may take $\Gamma = O^\times$. (The scalars do not play a role in the theory of embeddings: they act by the identity under conjugation.)

For $\gamma \in \Gamma$, and an optimal embedding $\phi \in \mathrm{Emb}(S, O)$, we obtain a new embedding via $\alpha \mapsto \gamma^{-1}\phi(\alpha)\gamma$, i.e., Γ acts on the *right* on $\mathrm{Emb}(S, O)$ by conjugation, and correspondingly on the right on E by right multiplication.

Let

$$\mathrm{Emb}(S, O; \Gamma) := \{\Gamma\text{-conjugacy classes of optimal } S \hookrightarrow O\} \tag{30.3.11}$$

and

$$m(S, O; \Gamma) := \# \operatorname{Emb}(S, O; \Gamma). \tag{30.3.12}$$

The quantity $m(S, O; \Gamma)$ only depends on the type (isomorphism class) of O (transporting Γ under the isomorphism of orders, of course).

By Lemma 30.3.8, there is a bijection

$$\operatorname{Emb}(S, O; \Gamma) \xrightarrow{\sim} K^\times \backslash E / \Gamma. \tag{30.3.13}$$

We conclude this section by a comparison: for groups Γ sitting between unit groups and norm 1 unit groups, we can compare embedding numbers as follows.

Lemma 30.3.14. *If $O^1 \leq \Gamma \leq O^\times$, then*

$$m(S, O; \Gamma) = m(S, O; O^\times)[\operatorname{nrd}(O^\times) : \operatorname{nrd}(\Gamma) \operatorname{nrd}(S^\times)]. \tag{30.3.15}$$

Proof. We have a surjective map

$$\operatorname{Emb}(S, O; \Gamma) \to \operatorname{Emb}(S, O; O^\times)$$

and the lemma amounts to looking at the fibers. From (30.3.13), we turn instead to

$$K^\times \backslash E / \Gamma \to K^\times \backslash E / O^\times.$$

For $\beta \in E$, the fiber over $K^\times \beta O^\times$ is

$$K^\times \backslash K^\times \beta O^\times / \Gamma \leftrightarrow K^{\beta\times} \backslash K^{\beta\times} O^\times / \Gamma \leftrightarrow \left(K^{\beta\times} \cap O^\times\right) \backslash O^\times / \Gamma \tag{30.3.16}$$

where $K^\beta = \beta^{-1} K \beta$. But by hypothesis on β, we have $K^{\beta\times} \cap O^\times = S^{\beta\times}$, so the fiber is in bijection with

$$K^\times \backslash K^\times \beta O^\times / \Gamma \leftrightarrow S^{\beta\times} \backslash O^\times / \Gamma.$$

Finally, the reduced norm gives a homomorphism $O^\times \to R^\times$ with kernel O^1, so since $O^1 \leq \Gamma \leq O^\times$, we have

$$\# S^{\beta\times} \backslash O^\times / \Gamma = \# \operatorname{nrd}(O^\times) / \operatorname{nrd}(S^{\beta\times}\Gamma) = [\operatorname{nrd}(O^\times) : \operatorname{nrd}(\Gamma) \operatorname{nrd}(S^\times)]$$

independent of β, giving the result. □

Remark 30.3.17. The term *optimal* goes back at least to Schilling [Schi35], but the notion was studied in the context of maximal orders as well by Chevalley [Chev34], Hasse [Hass34], and Noether [Noe34]. The theory of optimal embeddings was developed thereupon by Eichler [Eic56a, §3]; many key ideas can be seen transparently in Eichler [Eic73, Chapter II, §§3–5]. For further history up to the present, see Remark 30.6.18.

30.4 Counting embeddings, idelically: the trace formula

In this section, we give a formula the number of conjugacy classes of embeddings, using local-global (idelic) methods. We retain notation from the previous section, but now specialize to the case where R is a global ring with $F = \operatorname{Frac} R$. We use adelic (mostly idelic) notation as in 27.6.4.

We began in the previous section with a first embedding $K \hookrightarrow B$. As a reminder, the existence of such an embedding is determined by a local-global principle as follows.

30.4.1. By the local-global principle for embeddings (Proposition 14.6.7, and 14.6.8 for the case $K \simeq F \times F$), there exists an F-algebra embedding $K \hookrightarrow B$ if and only if there exist F_v-algebra embeddings $K_v \hookrightarrow B_v$ for all $v \in \operatorname{Pl} F$ if and only if K_v is a field for all $v \in \operatorname{Ram} B$.

30.4.2. The definitions in the previous section extend to each completion. Let $\widehat{K} = K \otimes_F \widehat{F}$ and similarly $\widehat{S} = S \otimes_R \widehat{R}$. Let

$$\widehat{\Gamma} = (\Gamma_v)_v \leq N_{\widehat{B}^\times} \widehat{O}$$

be a subgroup whose image in $N_{\widehat{B}^\times}(\widehat{O})/\widehat{F}^\times$ has finite index. For example, we can take $\widehat{\Gamma}$ the congruence closure of Γ.

We then analogously define

$$\operatorname{Emb}_{\widehat{R}}(\widehat{S}, \widehat{O}; \widehat{\Gamma}) := \{\widehat{\Gamma}\text{-conjugacy classes of optimal } \widehat{S} \hookrightarrow \widehat{O}\} \tag{30.4.3}$$

and

$$m(\widehat{S}, \widehat{O}; \widehat{\Gamma}) := \# \operatorname{Emb}_{\widehat{R}}(\widehat{S}, \widehat{O}; \widehat{\Gamma}). \tag{30.4.4}$$

30.4.5. As in Lemma 30.3.8, we define

$$\widehat{E} = E_{\widehat{S},\widehat{O}} := \left\{\widehat{\beta} \in \widehat{B}^\times : \widehat{\beta}^{-1}\widehat{K}\widehat{\beta} \cap \widehat{O} = \widehat{\beta}^{-1}\widehat{S}\widehat{\beta}\right\}$$

and obtain a bijection

$$\operatorname{Emb}(\widehat{S}, \widehat{O}; \widehat{\Gamma}) \xrightarrow{\sim} \widehat{K}^\times \backslash \widehat{E}/\widehat{\Gamma}. \tag{30.4.6}$$

under conjugation.

We now show that the number of global embeddings is counted by a class number times the number of local embeddings. In view of Lemma 30.3.14, we may focus on the case $\widehat{\Gamma} = \widehat{O}^\times$. As usual, we write $\operatorname{Cls} O$ for the right class set of O.

Theorem 30.4.7 (Trace formula). *Let $h(S) := \# \operatorname{Pic} S$. Then*

$$\sum_{[I] \in \operatorname{Cls} O} m(S, O_{\mathsf{L}}(I); O_{\mathsf{L}}(I)^\times) = h(S) m(\widehat{S}, \widehat{O}; \widehat{O}^\times). \tag{30.4.8}$$

See Remark 30.4.24 for an explanation of the name *trace formula*.

Proof. We decompose the set $K^\times\backslash\widehat{E}/\widehat{O}^\times$ in two different ways.

First, there is a natural map

$$K^\times\backslash\widehat{E}/\widehat{O}^\times \to \widehat{K}^\times\backslash\widehat{E}/\widehat{O}^\times \tag{30.4.9}$$

which is a surjective map of pointed sets. The fiber of (30.4.9) over the identity element is

$$K^\times\backslash\widehat{K}^\times/\widehat{S}^\times \simeq \operatorname{Pic} S. \tag{30.4.10}$$

We claim that the fibers of (30.4.9) may be similarly identified. Indeed, the fiber over $\widehat{K}^\times\widehat{\beta}\widehat{O}^\times$ consists of the double cosets $K^\times\widehat{\nu}\widehat{\beta}\widehat{O}^\times$ with $K^\times\widehat{\nu} \in K^\times\backslash\widehat{K}^\times$, and

$$K^\times\widehat{\nu}\widehat{\beta}\widehat{O}^\times = K^\times\widehat{\beta}\widehat{O}^\times$$

if and only if

$$\widehat{\nu}\widehat{\beta} = \rho\widehat{\beta}\widehat{\mu}$$

with $\rho \in K^\times$ and $\widehat{\mu} \in \widehat{O}^\times$, if and only if

$$\rho^{-1}\widehat{\nu} = \widehat{\beta}\widehat{\mu}\widehat{\beta}^{-1} \in \widehat{K}^\times \cap \widehat{\beta}\widehat{O}^\times\widehat{\beta}^{-1} = \widehat{S}^\times \tag{30.4.11}$$

(since $\widehat{\beta} \in \widehat{E}$), if and only if $K^\times\widehat{\nu} \subseteq K^\times\widehat{S}^\times$, as claimed. From the claim and (30.4.6), we conclude that

$$\#(K^\times\backslash\widehat{E}/\widehat{O}^\times) = h(S)m(\widehat{S}, \widehat{O}; \widehat{O}^\times). \tag{30.4.12}$$

On the other hand, each $\widehat{\beta}\widehat{O}^\times \in \widehat{E}/\widehat{O}^\times$ defines a right $\widehat{O}$-ideal; intersecting with B and organizing these right ideals by their classes, we will now show that they give rise to optimal embeddings of the corresponding left order. For brevity, write $O_I = O_{\mathsf{L}}(I)$. There is a map of pointed sets

$$K^\times\backslash\widehat{E}/\widehat{O}^\times \to B^\times\backslash\widehat{B}^\times/\widehat{O}^\times \xrightarrow{\sim} \operatorname{Cls} O. \tag{30.4.13}$$

Choose representatives

$$B^\times\backslash\widehat{B}^\times/\widehat{O}^\times = \bigsqcup_{[I]\in\operatorname{Cls} O} B^\times\widehat{\alpha}_I\widehat{O}^\times \tag{30.4.14}$$

so that $I = \widehat{\alpha}_I\widehat{O} \cap B$. Then

$$O_I = O_L(I) = \widehat{\alpha}_I\widehat{O}\widehat{\alpha}_I^{-1} \cap B.$$

Let

$$E_I = \{\beta \in B^\times : K \cap \beta O_I\beta^{-1} = S\}. \tag{30.4.15}$$

Now if

$$\widehat{\beta}\widehat{O}^\times \in \widehat{E}/\widehat{O}^\times,$$

then there exists a unique I such that

$$B^\times\widehat{\beta}\widehat{O}^\times \subseteq B^\times\widehat{\alpha}_I\widehat{O}^\times,$$

and therefore

$$\widehat{\beta}\widehat{O}^\times = (\beta\widehat{\alpha}_I)\widehat{O}^\times$$

for some $\beta \in B^\times$. If $\beta' \in B^\times$ is another, then

$$\beta^{-1}\beta' \in \widehat{\alpha}_I\widehat{O}^\times\widehat{\alpha}_I^{-1} \cap B = O_I^\times$$

so the class $\beta O_I^\times$ is well-defined.

We claim that $\beta \in E_I$, and conversely if $\beta \in E_I$ then $\widehat{\beta} = \beta\widehat{\alpha}_I^{-1} \in \widehat{E}$. Indeed,

$$\begin{aligned} K \cap \beta O_I \beta^{-1} &= K \cap (\widehat{\beta}\widehat{\alpha}_I^{-1})(\widehat{\alpha}_I\widehat{O}\widehat{\alpha}_I^{-1})(\widehat{\alpha}_I\widehat{\beta}^{-1}) \\ &= K \cap \widehat{\beta}\widehat{O}\widehat{\beta}^{-1} \end{aligned}$$

so $K \cap \beta O_I \beta^{-1} = S$ if and only if $\widehat{K} \cap \widehat{\beta}\widehat{O}\widehat{\beta}^{-1} = \widehat{S}$. Therefore there is a bijection

$$\begin{aligned} K^\times\backslash\widehat{E}/\widehat{O}^\times &\xrightarrow{\sim} \bigsqcup_I K^\times\backslash E_I/O_I^\times \xrightarrow{\sim} \bigsqcup_I \operatorname{Emb}(S, O_I; O_I^\times) \\ K^\times\widehat{\beta}\widehat{O}^\times &\mapsto K^\times\beta O_I^\times. \end{aligned} \tag{30.4.16}$$

Putting (30.4.12) and (30.4.16) together and counting, the theorem follows. □

When O has class number 1, we hit the embedding number on the nose.

Corollary 30.4.17. *If* $\#\operatorname{Cls} O = 1$, *then*

$$m(S, O; O^\times) = h(S)m(\widehat{S}, \widehat{O}; \widehat{O}^\times).$$

Proof. Immediate from Theorem 30.4.7. □

We recall that the genus of O (Definition 17.4.8) is the set $\operatorname{Gen} O$ of R-orders $O' \subseteq B$ locally isomorphic to O.

Corollary 30.4.18. *If* $\operatorname{Emb}(S_{\mathfrak{p}}, O_{\mathfrak{p}}) \neq \emptyset$ *for all primes* $\mathfrak{p}$, *then there exists an order* $O' \in \operatorname{Gen} O$ *such that* $\operatorname{Emb}(S, O') \neq \emptyset$.

Proof. By the trace formula, we have $\prod_{\mathfrak{p}} m(S_{\mathfrak{p}}, O_{\mathfrak{p}}; O_{\mathfrak{p}}^\times) = m(\widehat{S}, \widehat{O}; \widehat{O}^\times) > 0$, so since all terms in the formula are nonnegative integers, we have $m(S, O_{\mathsf{L}}(I); O_{\mathsf{L}}(I)^\times) > 0$ for some $[I] \in \operatorname{Cls} O$, and $O_{\mathsf{L}}(I) \in \operatorname{Gen} O$. □

Remark 30.4.19. The possible failure of local optimal embeddings to 'glue' to a global optimal embedding into O is measured by the phenomenon of *selectivity*, studied in Chapter 31.

30.4.20. More generally, an isomorphism $\phi\colon O \xrightarrow{\sim} O'$ of orders induces a bijection $\operatorname{Emb}_R(S, O) \leftrightarrow \operatorname{Emb}_R(S, O')$, and $\phi(O^\times) = O'^\times$, so we may group together the terms in (30.4.8) according to the type set $\operatorname{Typ} O$. Let

$$h(O) := [\operatorname{Idl} O : \operatorname{PIdl} O]$$

be index of the subgroup principal two-sided fractional O-ideals inside the invertible ones, studied in section 18.5. By Proposition 18.5.10, the equation (30.4.8) then becomes

$$\sum_{[O']\in \operatorname{Typ} O} h(O')m(S,O';O'^{\times}) = h(S)m(\widehat{S},\widehat{O};\widehat{O}^{\times}). \tag{30.4.21}$$

Remark 30.4.22. The foundational formula (30.4.8) is proven by counting a set two different ways. As such, it admits a purely combinatorial proof: Brzezinski [Brz89] shows that it follows from looking at "transitive actions of groups on pairs of sets and on relations invariant with respect to these actions" [Brz89, p. 199].

We also record the following corollary (in the vein of Remark 30.4.22).

Corollary 30.4.23. *Left multiplication defines a group action*

$$K^{\times}\backslash\widehat{K}^{\times}/\widehat{S}^{\times} \circlearrowright K^{\times}\backslash\widehat{E}/\widehat{O}^{\times}$$

that defines a free action of the group Pic S *on the set*

$$\bigsqcup_{[I]\in \operatorname{Cls} O} \operatorname{Emb}(S, O_{\mathsf{L}}(I); O_{\mathsf{L}}(I)^{\times}).$$

Proof. In the proof of the trace formula, we identified $K^{\times}\backslash\widehat{K}^{\times}/\widehat{S}^{\times} \simeq \operatorname{Pic} S$ and established an idelic to global bijection in (30.4.16); in (30.4.11), each fiber of the map (30.4.9) is identified with Pic S, describing the orbits of the action. (See also Exercise 30.12, where a direct proof is requested.) □

Remark 30.4.24. Theorem 30.4.7 is called a *trace formula* as it can be applied to compute the trace of certain matrices (called *Brandt matrices*) that encode the action of Hecke operators on a space of modular forms. We return to this important point of view in detail in Chapter 41, and apply the above formula to compute traces in section 41.5.

30.5 Local embedding numbers: maximal orders

In view of Theorem 30.4.7, we see that up to a class number of the base ring, optimal embeddings are counted in purely local terms. In this section and the next, we compute the relevant local embedding numbers; after that, we will return to the global setting to put the results together.

To this end, in this section and the next, we suppose R is local (as in 23.2.1): so R is a complete DVR with maximal ideal $\mathfrak{p} = \pi R$ and finite residue field $k := R/\mathfrak{p}$.

30.5.1. Since we are now local, the R-order S is free and so $S = R[\gamma]$ for some (not unique) $\gamma \in S$. Let $f_\gamma(x) := x^2 - t_\gamma x + n_\gamma$ be the minimal polynomial of γ, and let $d_\gamma := t_\gamma^2 - 4n_\gamma$ be the discriminant of f_γ, equal to the discriminant of S in $R/R^{\times 2}$. An R-algebra embedding from S is then determined uniquely by the image of γ.

We now compute (local) embedding numbers in the case of a maximal order; to do so, we introduce some notation.

30.5.2. Recalling that $K \supseteq F$ is a separable quadratic F-algebra and $\mathfrak{p}$ is the maximal ideal of R, we define a symbol recording the splitting of the prime $\mathfrak{p}$ in K (mirroring the Kronecker symbol):

$$\left(\frac{K}{\mathfrak{p}}\right) := \begin{cases} 1, & \text{if } K \simeq F \times F \text{ is split;} \\ 0, & \text{if } K \supseteq F \text{ is a ramified field extension;} \\ -1 & \text{if } K \supseteq F \text{ is an unramified field extension.} \end{cases}$$

If $q = \#k$ is odd, and $K \simeq F[x]/(x^2 - d)$, then

$$\left(\frac{K}{\mathfrak{p}}\right) = \left(\frac{d}{\mathfrak{p}}\right)$$

is the usual Legendre symbol.

Proposition 30.5.3. *The following statements hold.*

(a) *We have* $m(S, \mathrm{M}_2(R); \mathrm{GL}_2(R)) = 1$.
(b) *Suppose B is a division algebra and $O \subseteq B$ its valuation ring. If K is a field and $S = R_K$ is integrally closed, then*

$$m(S, O; N_{B^\times}(O)) = 1$$
$$m(S, O; O^\times) = 1 - \left(\frac{K}{\mathfrak{p}}\right).$$

If K is not a field or if S is not integrally closed, then $\mathrm{Emb}(S, O) = \emptyset$.

We recall that $N_{\mathrm{GL}_2(F)}(\mathrm{M}_2(R)) = F^\times \mathrm{GL}_2(R)$, so (a) also includes the case of the normalizer.

Proof. First, part (a). We have at least one embedding $\phi\colon S \hookrightarrow \mathrm{End}_R(S) \simeq \mathrm{M}_2(R)$ given by the regular representation of S on itself (in a basis): in the basis $1, \gamma$, we have

$$\gamma \mapsto \begin{pmatrix} 0 & -n_\gamma \\ 1 & t_\gamma \end{pmatrix} \tag{30.5.4}$$

a matrix in rational canonical form. This embedding is optimal, because if $x, y \in F$ satisfy

$$\phi(x + y\gamma) = \begin{pmatrix} x & -n_\gamma y \\ y & x + t_\gamma y \end{pmatrix} \in \mathrm{M}_2(R)$$

then $x, y \in R$ already, so $\phi(K) \cap \mathrm{M}_2(R) = \phi(S)$.

To finish (a), we need to show that the embedding (30.5.4) is the unique one, up to conjugation by $\mathrm{GL}_2(R)$. So let $\psi\colon S \hookrightarrow \mathrm{M}_2(R)$ be another optimal embedding. Then via ψ, the R-module $M = R^2$ (column vectors) has the structure of a left S-module; the condition that ψ is optimal translates into the condition that the (left) multiplicator ring of R^2 in K is precisely S, and therefore M is *invertible* as a left S-module by Exercise

16.13, and therefore principal, generated by $x \in M$, so that $M = Sx = Rx + R\gamma x$. In the R-basis $x, \gamma x$, the left regular representation has the form (30.5.4), completing the proof.

Here is a second quick matrix proof (finding explicitly the cyclic basis). Let $\gamma \mapsto \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in \mathrm{M}_2(R)$ be an optimal embedding; then at least one of $b, c, d - a \in R^\times$. Therefore there exist $x_1, x_2 \in R$ such that $q(x_1, x_2) = cx_1^2 + (d-a)x_1x_2 - bx_2^2 \in R^\times$. Let $x = (x_1, x_2)^{\mathsf{t}}$; then $\gamma x = (ax_1 + bx_2, cx_1 + dx_2)^{\mathsf{t}}$. Let $\alpha \in \mathrm{M}_2(R)$ be the matrix with columns $x, \gamma x$. Then $\det \alpha = q(x_1, x_2) \in R^\times$, so in fact $\alpha \in \mathrm{GL}_2(R)$. We then compute

$$\alpha^{-1} \begin{pmatrix} a & b \\ c & d \end{pmatrix} \alpha = \begin{pmatrix} 0 & -(ad-bc) \\ 1 & a+d \end{pmatrix}$$

as claimed. (This matrix proof will be generalized in the next section, when we work more generally with Eichler orders.)

Next, part (b). By Corollary 13.4.5, there exists an embedding $K \hookrightarrow B$ if and only if K is a field, so suppose K is a field. By Proposition 13.3.4, the valuation ring $\mathcal{O}$ is the unique maximal R-order in B, consisting of all integral elements, so the embedding restricts to an embedding $S \hookrightarrow \mathcal{O}$ by uniqueness. Suppose $S = R_K$; then an embedding $S \hookrightarrow \mathcal{O}$ extends to $K \hookrightarrow B$ so is conjugate 30.3.1 to every other under the action of $B^\times$. But the valuation ring is unique, thus $N_{B^\times}(\mathcal{O}) = B^\times$. Thus $m(S, \mathcal{O}; N_{B^\times}(\mathcal{O})) = 1$. Finally, we have $N_{B^\times}(\mathcal{O})/(F^\times \mathcal{O}^\times) \simeq \mathbb{Z}/2\mathbb{Z}$, generated by any $j \in \mathcal{O}$ with $\operatorname{nrd}(j) = \pi$. If the extension $K \supseteq F$ is ramified, then $K = F(j)$ so j centralizes K, and $m(S, \mathcal{O}; N_{B^\times}(\mathcal{O})) = m(S, \mathcal{O}; \mathcal{O}^\times) = 1$. If instead $K \supseteq F$ is inert, then $B \simeq (K, j \mid F)$, and conjugation by j normalizes but does not centralize K, and $m(S, \mathcal{O}; \mathcal{O}^\times) = 2$. ☐

In all cases, the local embedding numbers are finite.

Corollary 30.5.5. $m(S, \mathcal{O}; \Gamma) < \infty$.

Proof. Recalling that the image of Γ in $N_{B^\times}(\mathcal{O})/F^\times$ is a subgroup of finite index, we know that the natural surjective map

$$\operatorname{Emb}(S, \mathcal{O}; \Gamma) \to \operatorname{Emb}(S, \mathcal{O}; N_{B^\times}(\mathcal{O}))$$

is finite-to-one; applying this argument twice we reduce to the case that $\Gamma = \mathcal{O}^\times$.

Let $\mathcal{O}' \supseteq \mathcal{O}$ be a maximal R-order. Then $\mathcal{O}^\times \leq \mathcal{O}'^\times$. Moreover, each $\phi \in \operatorname{Emb}(S, \mathcal{O})$ gives (by composing with $\mathcal{O} \hookrightarrow \mathcal{O}'$) an embedding $\phi \colon S \hookrightarrow \mathcal{O}'$ that is optimal for some superorder $S' \supseteq S$. Thus

$$m(S, \mathcal{O}; \mathcal{O}^\times) \leq \sum_{S' \supseteq S} [\mathcal{O}'^\times : \mathcal{O}^\times] m(S', \mathcal{O}'; \mathcal{O}'^\times);$$

There are only finitely many superorders S' in the sum, since the integral closure R_K contains all S' so $[R_K : S']_R \mid [R_K : S]_R$—or equally well, compare discriminants. Applying Proposition 30.5.3, we conclude $m(S, \mathcal{O}; \mathcal{O}^\times) < \infty$. ☐

30.6 ∗ Local embedding numbers: Eichler orders

At this point, the presentation of local embedding numbers begins to run off the rails: for more general classes of orders, formulas for local embedding numbers are rarely as simple as in Proposition 30.5.3. To avoid tumbling too far, we present formulas for the case where O is a local Eichler (residually split) order following Hijikata [Hij74, Theorem 2.3]. See Remark 30.6.18 for further reference.

In this section, we retain the assumption that R is local and the notation in 30.5.1. Further, we suppose O is an Eichler order of level $\mathfrak{p}^e$ with $e \geq 0$.

30.6.1. Let

$$\varpi = \begin{pmatrix} 0 & 1 \\ \pi^e & 0 \end{pmatrix} \in O;$$

then

$$N_{B^\times}(O)/F^\times O^\times = \langle \varpi \rangle. \tag{30.6.2}$$

If $e = 0$, then $\varpi \in O^\times$ and $N_{B^\times}(O) = F^\times O^\times$; if instead $e \geq 1$, then $\langle \varpi \rangle \simeq \mathbb{Z}/2\mathbb{Z}$.

For $\phi \in \mathrm{Emb}(S, O)$, let ϕ^ϖ be defined by

$$\begin{aligned} \phi^\varpi &: S \hookrightarrow O \\ \phi^\varpi(\alpha) &= \varpi^{-1}\phi(\alpha)\varpi. \end{aligned}$$

By (30.6.2), $\phi, \phi' \in \mathrm{Emb}(S, O)$ are equivalent in $\mathrm{Emb}(S, O; N_{B^\times}(O))$ if and only if ϕ' is equivalent to either ϕ or ϕ^ϖ in $\mathrm{Emb}(S, O; O^\times)$.

Lemma 30.6.3. *Let $\phi \in \mathrm{Emb}(S, O)$. Then there exist $x \in R$ and $\nu \in N_{B^\times}(O)$ such that*

$$f_\gamma(x) = x^2 - t_\gamma x + n_\gamma \equiv 0 \pmod{\mathfrak{p}^e}$$

and

$$\nu^{-1}\phi(\gamma)\nu = \begin{pmatrix} x & 1 \\ -f_\gamma(x) & t_\gamma - x \end{pmatrix}. \tag{30.6.4}$$

Proof. We may suppose that O is the standard Eichler order. Let $\phi \colon S \hookrightarrow O$ be an embedding, with

$$\phi(\gamma) = \begin{pmatrix} a & b \\ c\pi^e & d \end{pmatrix} \in O$$

so that $a, b, c, d \in R$. We have $t_\gamma = \mathrm{trd}(\gamma) = a + d$ and $n_\gamma = \mathrm{nrd}(\gamma) = ad - bc\pi^e$. We observe that ϕ is optimal if and only if at least one of $b, c, a - d \in R^\times$: indeed, there exists $z \in R$ such that $(\gamma - z)/\pi \in O$ if and only if all belong to $\mathfrak{p}$.

Suppose now that ϕ is optimal. We have three cases.

If $b \in R^\times$, then take $\nu = \begin{pmatrix} 1 & 0 \\ 0 & b^{-1} \end{pmatrix} \in O^\times$; we compute

$$f_\gamma(a) = a^2 - (a + d)a + (ad - bc\pi^e) = -bc\pi^e$$

and

$$\nu^{-1}\phi(\gamma)\nu = \begin{pmatrix} a & 1 \\ bc\pi^e & d \end{pmatrix} = \begin{pmatrix} a & 1 \\ -f_\gamma(a) & t_\gamma - a \end{pmatrix} \tag{30.6.5}$$

as desired.

If $c \in R^\times$, then we take $\nu = \varpi = \begin{pmatrix} 0 & 1 \\ \pi^e & 0 \end{pmatrix}$; now

$$\nu^{-1}\phi(\gamma)\nu = \begin{pmatrix} d & c \\ b\pi^e & a \end{pmatrix} \tag{30.6.6}$$

and we apply the previous case.

Finally, if $a - d \in R^\times$, we may suppose $b \in \mathfrak{p}$ as well as $c \in \mathfrak{p}$ if $e = 0$, and then we take $\nu = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}$ to find

$$\nu^{-1}\phi(\gamma)\nu = \begin{pmatrix} a - c\pi^e & a - d + b - c\pi^e \\ c\pi^e & c\pi^e + d \end{pmatrix} \tag{30.6.7}$$

to reduce again to the first case. □

Definition 30.6.8. An optimal embedding $\phi \in \mathrm{Emb}(S, O)$ is **normalized** and **associated** to $x \in R$ if

$$\phi(\gamma) = \begin{pmatrix} x & 1 \\ -f_\gamma(x) & t_\gamma - x \end{pmatrix}$$

(as in (30.6.4)).

The statement of Lemma 30.6.3 is that for all $\phi \in \mathrm{Emb}(S, O)$, either the class of ϕ or of ϕ^ϖ in $\mathrm{Emb}(S, O; O^\times)$ is represented by a normalized embedding. To conclude our efforts, we need to check which of these are equivalent. We recall that $d_\gamma = t_\gamma^2 - 4n_\gamma$.

Lemma 30.6.9. *Let ϕ, ϕ' be normalized embeddings associated to $x, x' \in R$. Then the following statements hold.*

(a) *ϕ, ϕ' are conjugate by $O^\times$ if and only if $x \equiv x' \pmod{\mathfrak{p}^e}$.*
(b) *If $d_\gamma \in R^\times$ or $e = 0$, then ϕ^ϖ is equivalent to ϕ' in $\mathrm{Emb}(S, O; O^\times)$ if and only if $x' \equiv t_\gamma - x \pmod{\mathfrak{p}^e}$.*
(c) *If $d_\gamma \notin R^\times$ and $e \geq 1$, then ϕ^ϖ is equivalent to ϕ' in $\mathrm{Emb}(S, O; O^\times)$ if and only if $x' \equiv t_\gamma - x \pmod{\mathfrak{p}^e}$ and $f_\gamma(x) \not\equiv 0 \pmod{\mathfrak{p}^{e+1}}$.*

Proof. First (a). If ϕ, ϕ' are conjugate by some $\mu \in O^\times$, we reduce modulo $\mathfrak{p}^e\,\mathrm{M}_2(R)$ to obtain the ring of upper triangular matrices, and see that the diagonal entries of $\phi(\gamma), \phi'(\gamma)$ are congruent modulo $\mathfrak{p}^e$ and in particular $x \equiv x' \pmod{\mathfrak{p}^e}$. Conversely, if $x \equiv x' \pmod{\mathfrak{p}^e}$, then let $\mu = \begin{pmatrix} 1 & 0 \\ x' - x & 1 \end{pmatrix}$; we confirm that

$$\mu^{-1}\phi(\gamma)\mu = \begin{pmatrix} x' & 1 \\ -f_\gamma(x') & t_\gamma - x' \end{pmatrix} = \phi'(\gamma).$$

In preparation for (b) and (c), we note that

$$\phi^\varpi(\gamma) = \begin{pmatrix} t_\gamma - x & -f_\gamma(x)/\pi^e \\ \pi^e & x \end{pmatrix}. \tag{30.6.10}$$

We now prove (b). If $e = 0$, the statement is true: $\varpi \in O^\times$ and all embeddings are conjugate. Suppose $e \geq 1$. Since $f_\gamma(x) \equiv 0 \pmod{\mathfrak{p}^e}$ we have

$$d_\gamma \equiv (x - \overline{x})^2 = (x - (t_\gamma - x))^2 = (t_\gamma - 2x)^2 \pmod{\mathfrak{p}^e}$$

so $t_\gamma - 2x \in R^\times$. For $u \in R$, let $\mu = \begin{pmatrix} 1 & u \\ 0 & 1 \end{pmatrix}$ and consider

$$\begin{aligned} \mu^{-1}\phi^\varpi(\gamma)\mu &= \begin{pmatrix} t_\gamma - x - u\pi^e & u(t_\gamma - 2x) - u^2\pi^e - \pi^{-e} f_\gamma(x) \\ \pi^e & x + u\pi^e \end{pmatrix} \\ &\equiv \begin{pmatrix} t_\gamma - x & u(t_\gamma - 2x) - f_\gamma(x)/\pi^e \\ \pi^e & x \end{pmatrix} \pmod{\mathfrak{p}^e}. \end{aligned} \tag{30.6.11}$$

Thus we may choose u so that the top-right entry of (30.6.11) is equal to 1. The result then follows from (a).

We conclude with (c), and we are given $e \geq 1$ and $d_\gamma \notin R^\times$. If $\pi^{-e} f_\gamma(x) \in R^\times$ then from (30.6.10) and looking back at (30.6.5), already ϕ^ϖ is conjugate under $O^\times$ to the normalized embedding associated to $t_\gamma - x$, so by (a), we then have ϕ^ϖ equivalent to ϕ' if (and only if) $x \equiv t - x \pmod{\mathfrak{p}^e}$. To finish, suppose $f_\gamma(x) \in \mathfrak{p}^{e+1}$ and $d \in \mathfrak{p}$. Reducing modulo $\mathfrak{p}^e\,\mathrm{M}_2(R)$, the ring $O/(\mathfrak{p}^e\,\mathrm{M}_2(R))$ consists of upper-triangular matrices and its unit group is generated by diagonal matrices and units μ as in the previous paragraph. But then by (30.6.11), the top-right entry of every $O^\times$-conjugate of $\phi^\varpi(\gamma)$ belongs to $\mathfrak{p}$, and so it cannot be equal to $\phi'(\gamma)$. □

The statements above give a way to compute the local embedding number in terms of arithmetic of R.

Proposition 30.6.12. *For $s \in \mathbb{Z}_{\geq 1}$, let*

$$M(s) := \{x \in R/\mathfrak{p}^s : f_\gamma(x) \equiv 0 \pmod{\mathfrak{p}^s}\}. \tag{30.6.13}$$

Then for $e \geq 1$,

$$m(S, O; O^\times) = \begin{cases} \#M(e), & \text{if } d_\gamma \in R^\times; \\ \#M(e) + \#\operatorname{img}(M(e+1) \to R/\mathfrak{p}^e), & \text{otherwise.} \end{cases}$$

Proof. By Lemma 30.6.3, the set $\operatorname{Emb}(S, O; O^\times)$ is represented by the set of normalized embeddings and their conjugates under ϖ. By Lemma 30.6.9(a), the normalized embeddings according to $x \in M(s)$ are distinct; by (b)–(c), the remaining conjugate embeddings are new when they lift to $M(e+1)$. □

Example 30.6.14. Let $R = \mathbb{Z}_2$ and $S = \mathbb{Z}_2[\sqrt{-1}]$, so $\gamma = \sqrt{-1}$ and $f_\gamma(x) = x^2 + 1$. We have

$$M(s) = \begin{cases} \{1 \bmod 2\}, & \text{if } s = 1; \\ \emptyset, & \text{if } s \geq 2. \end{cases}$$

Therefore by Proposition 30.6.12, if O is an Eichler order of level 2^e over $\mathbb{Z}_2$, then

$$m(\mathbb{Z}_2[\sqrt{-1}], O; O^\times) = \begin{cases} 1, & \text{if } e \leq 1; \\ 0, & \text{if } e \geq 2. \end{cases} \tag{30.6.15}$$

We record an important special case.

Lemma 30.6.16. *If $e = 1$ and S is a maximal R-order in K (equivalently, S is integrally closed in K), then*

$$m(S, O; O^\times) = 1 + \left(\frac{K}{\mathfrak{p}}\right).$$

Proof. Referring to Proposition 30.6.12, we first compute $\#M(1)$, the number of solutions to $f_\gamma(x) \equiv 0 \pmod{\mathfrak{p}}$, so over the finite field $k = R/\mathfrak{p}$: this number is $2, 1, 0$ according as $\left(\frac{K}{\mathfrak{p}}\right) = 1, 0, -1$. If $d \in R^\times$, we are done; so suppose $d \notin R^\times$, hence $\left(\frac{K}{\mathfrak{p}}\right) = 0$. Let $x \in M(2)$. We claim that $(\gamma - x)/\pi$ is integral over R, contradicting that S is maximal. Indeed, $\pi^2 \mid f_\gamma(x) = x^2 - t_\gamma x + n_\gamma = \operatorname{nrd}(\gamma - x)$; and $\mathfrak{p} \mid d_\gamma = t_\gamma^2 - 4n_\gamma = (t_\gamma - 2x)^2$, so $\mathfrak{p} \mid (t_\gamma - 2x) = \operatorname{trd}(\gamma - x)$. Thus $M(2) = \emptyset$ and so the count remains the same. □

To conclude, when $q = \#k$ is odd, we can make Proposition 30.6.12 completely explicit.

Lemma 30.6.17. *Suppose $\#k = q$ is odd, that $e \geq 1$, and let $f := \operatorname{ord}_{\mathfrak{p}}(d_\gamma)$.*

(a) *If $f = 0$, then*

$$m(S, O; O^\times) = 1 + \left(\frac{K}{\mathfrak{p}}\right).$$

(b) *If $e < f$, then*

$$m(S, O; O^\times) = \begin{cases} 2q^{(e-1)/2}, & \text{if } e \text{ is odd;} \\ q^{e/2-1}(q+1), & \text{if } e \text{ is even.} \end{cases}$$

(c) *If $e = f$, then*

$$m(S, O; O^\times) = \begin{cases} q^{(f-1)/2}, & \text{if } f \text{ is odd;} \\ q^{f/2} + q^{f/2-1}\left(1 + \left(\frac{K}{\mathfrak{p}}\right)\right), & \text{if } f \text{ is even.} \end{cases}$$

(d) *If $e > f > 0$, then*

$$m(S, O; O^\times) = \begin{cases} 0, & \text{if } f \text{ is odd;} \\ q^{f/2-1}(q+1)\left(1 + \left(\frac{K}{\mathfrak{p}}\right)\right), & \text{if } f \text{ is even.} \end{cases}$$

Proof. Since the residue field k has odd characteristic, we can complete the square and without loss of generality we may suppose that $\operatorname{trd}(\gamma) = 0$, and

$$M(s) = \left\{x \in R/\mathfrak{p}^s : x^2 \equiv d \pmod{\mathfrak{p}^s}\right\}.$$

We will abbreviate $m = m(S, O; O^\times)$ and repeatedly refer to Proposition 30.6.12.

First suppose $f = 0$. Then $d_\gamma \in R^\times$, so by Proposition 30.6.12, we have $m = \#M(e)$. But by Hensel's lemma, $\#M(e) = 0$ or 2 according as d is a square or not in R for all $e \geq 1$.

Next suppose that $e < f$. The solutions to the equation $x^2 \equiv 0 \pmod{\mathfrak{p}^s}$ are those with $x \equiv 0 \pmod{\mathfrak{p}^{\lceil s/2 \rceil}}$. Thus $\#M(e) = q^{e-\lceil e/2 \rceil} = q^{\lfloor e/2 \rfloor}$ and we see that $\#\operatorname{img}(M(e+1) \to R/\mathfrak{p}^e) = q^{e-\lceil (e+1)/2 \rceil}$, so $m = 2q^{(e-1)/2}$ if e is odd and $m = q^{e/2} + q^{e/2-1} = q^{e/2-1}(q+1)$ if e is even.

If $e = f$, then again $\#M(e) = q^{\lfloor e/2 \rfloor}$. To count the second contributing set, we must solve $x^2 \equiv d_\gamma \pmod{\mathfrak{p}^{e+1}}$. If $e = f$ is odd then this congruence has no solution. If instead e is even then we must solve $y^2 = (x/\pi^{f/2})^2 \equiv d_\gamma/\pi^f \pmod{\mathfrak{p}}$ where π is a uniformizer at $\mathfrak{p}$. This latter congruence has zero or two solutions according as d is a square, and given such a solution y we have the solutions $x \equiv y \pmod{\pi^{f/2+1}}$ to the original congruence, and hence there are 0 or $2q^{f-(f/2+1)} = 2q^{f/2-1}$ solutions, as claimed.

Finally, suppose $e > f > 0$. If f is odd, there are no solutions to $x^2 \equiv d_\gamma \pmod{\mathfrak{p}^e}$. If f is even, there are no solutions if d_γ is not a square and otherwise the solutions are $x \equiv y \pmod{\mathfrak{p}^{e-f/2}}$ as above so they total $2q^{f/2} + 2q^{f/2-1} = 2q^{f/2-1}(q+1)$. □

On the other hand, when the residue field k has even characteristic, the computations become even more involved!

Remark 30.6.18. Eichler studied optimal embeddings [Eic38b, §2] very early on, computing the contribution of units (coming from embeddings of $\mathbb{Z}[\omega]$ and $\mathbb{Z}[i]$ in a maximal order O) in the mass formula. He then [Eic56a, §3] studied more generally optimal embeddings of quadratic orders into his Eichler orders of squarefree level. Hijikata [Hij74, §2] studied optimal embeddings in the context of computing traces of Hecke operators on $\Gamma_0(N)$ (general N), with embedding numbers given for certain orders. See also Eichler's treatment [Eic73, §3] in the context of the basis problem for modular forms, as well as Pizer's presentation [Piz76a, §3]. (See Remark 41.5.13 for further detail.) Brzezinski [Brz91, Corollary 1.16] (a typo has it appear as Corollary 1.6) gives a general formula for Eichler orders (which is to say, a generalization of Lemma 30.6.17 to include q even)—the proof method is different than the method of Hijikata above, and the answer is organized a bit differently than Lemma 30.6.17.

But these papers are just the beginning, and there is a cornucopia of further results. Many of these are obtained in pursuit of progressively more general forms of the trace formula (see for example the summary of results by Hashimoto [Hash77]) for Eichler orders. Shimizu [Shz63, §§26–27] considered embedding numbers over totally real fields in computing the contribution of elliptic elements to formulas for the dimension of spaces of cusp forms and later for the trace formula [Shz65, §3]. The contributions of elliptic elements over totally real fields was also pursued by Prestel [Pre68, §5] and more generally for embeddings by Schneider [Schn75] (and quite explicitly for real quadratic fields [Schn77]) and Vignéras [Vig76a, §4].

Pizer [Piz76b, §§3–5] considered optimal embeddings for residually split orders (see 24.3.7) over $\mathbb{Q}$: these were then applied to further cases of the basis problem for modular forms [Piz76c, Piz80b]. Then Hijikata–Pizer–Shemanske [HPS89b, §§1–5] developed in a uniform manner the optimal embedding theory for basic orders (they

called them *special*, cf. Remark 24.5.7): the application to the trace formula is then contained in their monograph [HPS89a]. Brzezinski [Brz90, §3] also obtains recursive formulas for optimal embedding numbers of a local Bass (equivalently, basic) order (in characteristic not 2), using an effective description of the automorphism group of the order.

30.7 Global embedding numbers

We now combine the ingredients from the previous three sections to arrive at a formula for global embedding numbers.

We return to the setting of section 30.4, with R a global ring.

30.7.1. Our global ring $R = R_{(T)}$ comes from an eligible set $T \subseteq \operatorname{Pl} F$. (This set is usually denoted S, but we do not want any confusion with the quadratic R-algebra $S \subseteq K$.)

For all but finitely many places $v \notin T$, we have $O_v \simeq \mathrm{M}_2(R_v)$ maximal and $\Gamma_v = O_v^\times$. By Proposition 30.5.3(a), for such places v, we have $\#\operatorname{Emb}_{R_v}(S_v, O_v; O_v^\times) = 1$. Therefore the number $m(\widehat{S}, \widehat{O}; \widehat{\Gamma}) = \#\operatorname{Emb}_{\widehat{R}}(\widehat{S}, \widehat{O}; \widehat{\Gamma})$ of adelic embeddings is given by the (well-defined, finite) product

$$m(\widehat{S}, \widehat{O}; \widehat{\Gamma}) = \prod_{v \notin T} m(S_v, O_v; \Gamma_v) \tag{30.7.2}$$

(well-defined and finite).

We arrive at the following first global result.

Theorem 30.7.3. *Let* $\mathfrak{N} = \operatorname{discrd}(O)$. *Then*

$$\sum_{[I] \in \operatorname{Cls} O} m(S, O_{\mathsf{L}}(I); O_{\mathsf{L}}(I)^\times) = h(S) \prod_{\mathfrak{p} \mid \mathfrak{N}} m(S_\mathfrak{p}, O_\mathfrak{p}; O_\mathfrak{p}^\times).$$

Proof. For all $\mathfrak{p} \nmid \mathfrak{N}$, we have $O_\mathfrak{p} \simeq \mathrm{M}_2(R_\mathfrak{p})$, so the result follows by combining Theorem 30.4.7 with 30.7.1. ☐

Working with the normalizer instead, we find the following result.

Proposition 30.7.4. *Suppose that* O *is a maximal order with* $\#\operatorname{Cls} O = 1$. *Then*

$$m(S, O; N_{B^\times}(O)) = h(S).$$

Proof. As in the proof of Theorem 30.7.3, but with a further appeal to Proposition 30.5.3 to show that all local embeddings numbers are equal to 1. ☐

An important illustrative special case is the following.

Example 30.7.5. Let $\mathfrak{D} = \operatorname{disc}_R B$ and suppose that $O \subseteq B$ is an Eichler R-order of squarefree level $\mathfrak{M}$, so $\operatorname{discrd} O = \mathfrak{D}\mathfrak{M}$. Suppose further that S is a maximal R-order in K. Then Theorem 30.7.3 reads

$$\sum_{[I] \in \operatorname{Cls} O} m(S, O_{\mathsf{L}}(I); O_{\mathsf{L}}(I)^\times) = h(S) \prod_{\mathfrak{p} \mid \mathfrak{D}} \left(1 - \left(\frac{K}{\mathfrak{p}}\right)\right) \prod_{\mathfrak{p} \mid \mathfrak{M}} \left(1 + \left(\frac{K}{\mathfrak{p}}\right)\right),$$

with the local embedding numbers computed in Proposition 30.5.3(b) for $\mathfrak{p} \mid \mathfrak{D}$ and Lemma 30.6.16 for $\mathfrak{p} \mid \mathfrak{M}$.

Suppose further that B is T-indefinite and that $\#\operatorname{Cl}_\Omega R = 1$; then $\#\operatorname{Cls} O = \#\operatorname{Cl}_\Omega R$ by Corollary 28.5.17 (an application of strong approximation), so

$$m(S, O; O^\times) = h(S) \prod_{\mathfrak{p} \mid \mathfrak{D}} \left(1 - \left(\frac{K}{\mathfrak{p}}\right)\right) \prod_{\mathfrak{p} \mid \mathfrak{M}} \left(1 + \left(\frac{K}{\mathfrak{p}}\right)\right). \tag{30.7.6}$$

Embeddings of cyclotomic orders are of particular interest.

Example 30.7.7. Consider the case $F = \mathbb{Q}$ and $R = \mathbb{Z}$. Let $D = \operatorname{disc} B$ and suppose that O is an Eichler order of level M, so $N = DM = \operatorname{discrd} O$. We recall D is squarefree and $\gcd(D, M) = 1$. Suppose B is indefinite. Then by Theorem 28.2.11 (an application of strong approximation), $\#\operatorname{Cls} O = 1$.

If $K \supseteq \mathbb{Q}$ is a cyclotomic quadratic extension, then either $K = \mathbb{Q}(\omega) = \mathbb{Q}(\sqrt{-3})$ or $K = \mathbb{Q}(i) = \mathbb{Q}(\sqrt{-4})$ with corresponding rings of integers $\mathbb{Z}_K = \mathbb{Z}[\omega]$ and $\mathbb{Z}_K = \mathbb{Z}[i]$, each with $h(\mathbb{Z}_K) = 1$. With local embedding numbers computed in Lemma 30.6.17, Theorem 30.7.3 then gives

$$m(\mathbb{Z}[\omega], O; O^\times) = \begin{cases} \displaystyle\prod_{p \mid D} \left(1 - \left(\frac{-3}{p}\right)\right) \prod_{p \mid M} \left(1 + \left(\frac{-3}{p}\right)\right), & \text{if } 9 \nmid M; \\ 0, & \text{if } 9 \mid M. \end{cases}$$

Similarly, using (30.6.15),

$$m(\mathbb{Z}[i], O; O^\times) = \begin{cases} \displaystyle\prod_{p \mid D} \left(1 - \left(\frac{-4}{p}\right)\right) \prod_{p \mid M} \left(1 + \left(\frac{-4}{p}\right)\right), & \text{if } 4 \nmid M; \\ 0, & \text{if } 4 \mid M. \end{cases}$$

Absent further hypothesis, it is difficult to tease apart the term $m(S, O; O^\times)$ from the sum over left orders in Theorem 30.4.7. In the next chapter, we will show that the hypothesis that B is T-indefinite is *almost* enough.

30.8 Class number formula

In this section, we explain how the theory of optimal embeddings can be used to convert the mass formula into a class number formula, following Eichler.

Suppose throughout this section that B is T-definite. By Lemma 26.5.1, the group $O^\times/R^\times$ is finite; let $w_O = [O^\times : R^\times]$.

30.8.1. To a nontrivial cyclic subgroup of $O^\times/R^\times$, we associate the quadratic field K it generates over F; such a field is necessarily cyclotomic, with $K \simeq F(\zeta_{2q})$ for q the order of the cyclic subgroup. (Only certain $F(\zeta_{2q})$ arise as quadratic extensions of F, and different values of q can give rise to the same quadratic field up to isomorphism.)

Conversely, to a quadratic field $K \supseteq F$, we obtain a (possibly trivial) cyclic subgroup $(K^\times \cap O^\times)/R^\times$.

Lemma 30.8.2. *Every nontrivial $\alpha R^\times \in O^\times/R^\times$ belongs to a unique maximal cyclic subgroup.*

Proof. Since $O^\times/R^\times$ is finite, α belongs to at least one maximal cyclic subgroup; if it belonged to two, then the corresponding quadratic fields would both contain the field corresponding to α, hence by degrees would be equal, so by maximality the cyclic subgroups would have to be equal. □

30.8.3. Recall that $m(S,O;O^\times)$ counts optimal embeddings $\phi\colon S \hookrightarrow O$ up to conjugation by $O^\times$. Since $O^\times/R^\times$ is finite, the set $\operatorname{Emb}(S,O)$ is itself finite. Precisely two embeddings give rise to the same image $\phi(S)$, differing by the (necessarily nontrivial) standard involution. The stabilizer of $O^\times$ on $\phi \in \operatorname{Emb}(S,O)$ is $O^\times \cap \phi(S) = \phi(S)^\times \simeq S^\times$. Let $w_S = [S^\times : R^\times]$. We have shown that

$$m(S,O;O^\times) = \#\{\phi(S) \subseteq O : \phi \in \operatorname{Emb}(S;O)\}\frac{2w_S}{w_O}. \tag{30.8.4}$$

Proposition 30.8.5. *We have*

$$1 - \frac{1}{w_O} = \frac{1}{2}\sum_{q\geq 2}\left(1-\frac{1}{q}\right)\sum_{[S^\times:R^\times]=q} m(S,O;O^\times).$$

Proof. We count off the group $O^\times/R^\times$ by maximal cyclic subgroups, keeping track of the trivial class. By Lemma 30.8.2, every nontrivial $\alpha R^\times \in O^\times/R^\times$ belongs to a unique maximal cyclic subgroup of some order $q \geq 2$: such a subgroup is of the form $\phi(S)^\times/R^\times$ with $\phi(S) \subseteq O$ an optimally embedded order, and has $q-1$ nontrivial elements. Therefore

$$w_O - 1 = \sum_{q\geq 2}\sum_{\substack{S\subseteq K_q \\ [S^\times:R^\times]=q}} (q-1)\#\{\phi(S) \subseteq O : \phi \in \operatorname{Emb}(S;O)\}.$$

Plugging in (30.8.4), we obtain

$$w_O - 1 = \sum_{q\geq 2}\sum_{\substack{S\subseteq K_q \\ [S^\times:R^\times]=q}} (q-1)m(S,O;O^\times)\frac{w_O}{2q};$$

dividing through by w_O gives the result. □

We now recall the Eichler mass formula (Main Theorem 26.1.5), giving an explicit formula for the weighted class number

$$\operatorname{mass}(\operatorname{Cls} O) := \sum_{[I]\in \operatorname{Cls} O} \frac{1}{w_I},$$

where $w_I = [O_{\mathsf{L}}(I)^\times : R^\times]$.

Main Theorem 30.8.6 (Eichler class number formula). *Let R be a global ring with eligible set T, let B be an T-definite quaternion algebra over $F = \operatorname{Frac} R$, and let $O \subset B$ be an R-order. Then*

$$\#\operatorname{Cls} O = \operatorname{mass}(\operatorname{Cls} O) + \frac{1}{2}\sum_{q\geq 2}\left(1-\frac{1}{q}\right)\sum_{[S^\times:R^\times]=q} h(S)m(\widehat{S},\widehat{O};\widehat{O}^\times)$$

where the inner sum is over all quadratic R-algebras $S \supseteq R$ such that $[S^\times : R^\times] = q \in \mathbb{Z}_{\geq 2}$, and $h(S) = \#\operatorname{Pic} S$.

Proof. We apply Proposition 30.8.5 to each order $O_{\mathsf{L}}(I)$ for $[I] \in \operatorname{Cls} O$ and sum. We obtain

$$\begin{aligned}\sum_{[I]\in\operatorname{Cls} O}\left(1-\frac{1}{w_I}\right) &= \#\operatorname{Cls} O - \operatorname{mass}(\operatorname{Cls} O)\\ &= \frac{1}{2}\sum_{q\geq 2}\left(1-\frac{1}{q}\right)\sum_{[S^\times:R^\times]=q}\ \sum_{[I]\in\operatorname{Cls} O} m(S, O_{\mathsf{L}}(I); O_{\mathsf{L}}(I)^\times)\\ &= \frac{1}{2}\sum_{q\geq 2}\left(1-\frac{1}{q}\right)\sum_{[S^\times:R^\times]=q} h(S)m(\widehat{S},\widehat{O};\widehat{O}^\times),\end{aligned}$$

the last equality by Theorem 30.4.7. ◻

The expressions in the Eichler class number formula (Main Theorem 30.8.6) are arithmetically involved but can be computed effectively. In special cases, they give reasonably simple formulas, as previously advertised.

Example 30.8.7. When $F = \mathbb{Q}$, Main Theorem 30.8.6 yields the formula given in Theorem 30.1.5, with the computation of local class numbers given in Example 30.7.7.

Remark 30.8.8. The formula for the class number for Eichler orders of squarefree level (i.e., hereditary orders) was given by Eichler [Eic56a, Satz 10–11]; for Eichler orders over $\mathbb{Q}$ of arbitrary level (as in Example 30.8.7), it was given by Pizer [Piz76a, Theorem 16]. Main Theorem 30.8.6 is proven by Vignéras [Vig80a, Corollaire V.2.5] and Körner [Kör87, Theorem 2]. For further reference and discussion (in the context of computing local embedding numbers), see Remark 30.6.18.

30.8.9. Let $K \supset F$ be a quadratic field extension, let R_K be the integral closure of R in K, and let $S \subseteq K$ be an R-order of conductor $\mathfrak{f}$. Then there is an exact sequence

$$1 \to S^\times \to R_K^\times \to \frac{(R_K/\mathfrak{f}R_K)^\times}{(R/\mathfrak{f})^\times} \to \operatorname{Pic} S \to \operatorname{Pic} R_K \to 1 \tag{30.8.10}$$

giving rise to the formula of Dedekind

$$h(S) = \frac{h(R_K)}{[R_K^\times : S^\times]}\mathsf{N}(\mathfrak{f})\prod_{\mathfrak{p}\mid\mathfrak{f}}\left(1-\left(\frac{K}{\mathfrak{p}}\right)\frac{1}{\mathsf{N}(\mathfrak{p})}\right) \tag{30.8.11}$$

where N is the absolute norm and $\left(\dfrac{K}{\mathfrak{p}}\right)$ is given (globally) as in 30.5.2:

$$\left(\frac{K}{\mathfrak{p}}\right) = \begin{cases} -1, & \text{if } \mathfrak{p} \text{ is inert in } K; \\ 0, & \text{if } \mathfrak{p} \text{ is ramified in } K; \\ 1, & \text{if } \mathfrak{p} \text{ splits in } K. \end{cases}$$

30.9 Type number formula

We continue with the hypotheses of the previous section. A further application of the strategy to compute the class number is to also compute the type number. The methods are indeed quite similar: rearranging Corollary 18.5.12, we have

$$\# \operatorname{Typ} O = \frac{\# \operatorname{Cls} O}{\# \operatorname{Pic}_R O} + \sum_{[O'] \in \operatorname{Typ} O} \left(1 - \frac{1}{z_{O'}}\right) \tag{30.9.1}$$

where $z_{O'} = [N_{B^\times}(O') : F^\times O'^\times]$. But now the structure of the normalizer groups come into play, and one can give a type number formula similar to the class number formula 30.8.6 in terms of certain embedding numbers at least for Eichler orders. Unfortunately, even over $\mathbb{Q}$, these formulas quickly get very complicated! To give a sense of what can be proven, in this section we provide a type number formula in a special but interesting case due originally to Deuring [Deu51], and we refer to Remark 30.9.12 for further reference.

Proposition 30.9.2 (Deuring). *Let B be a definite quaternion algebra over $\mathbb{Q}$ with* $\operatorname{disc} B = p$ *prime and let $O \subset B$ be a maximal order. Then* $\# \operatorname{Typ} O = 1$ *for $p = 2, 3$, and for $p \geq 5$,*

$$\# \operatorname{Typ} O = \frac{1}{2} \# \operatorname{Cls} O + \frac{1}{4}\big([h(-p)] + h(-4p)\big)$$

where $[h(-p)] = h(-p)$ when $p \equiv 3 \pmod 4$ and is 0 otherwise.

Proof. In light of (30.9.1), we begin by considering the Picard group $\operatorname{Pic} O$ (with $R = \mathbb{Z}$): by 18.4.8, we have an isomorphism $\operatorname{Pic}(O) \simeq \mathbb{Z}/2\mathbb{Z}$ generated by the unique right ideal $J \subseteq O$ with $\operatorname{nrd}(J) = p$. The ideal J is automatically two-sided and contains all elements of reduced norm divisible by p (see 13.3.7); and thus J is principal if and only if there exists an element $\alpha \in O$ with $\operatorname{nrd}(\alpha) = p$ if and only if $z_O = 2$.

Therefore, (30.9.1) reads

$$\# \operatorname{Typ} O = \frac{1}{2} \# \operatorname{Cls} O + \frac{1}{2} \sum_{[O'] \in \operatorname{Typ} O} \#\{J' \subseteq O' \text{ principal right ideal} : \operatorname{nrd}(J') = p\}. \tag{30.9.3}$$

We now compute this sum in terms of embedding numbers. First, the map $\alpha \mapsto \alpha O$ gives

$$\begin{aligned} &\#\{J \subseteq O \text{ principal right ideal} : \operatorname{nrd}(J) = p\} \\ &\qquad = \frac{1}{2 w_O} \#\{\alpha \in O : \operatorname{nrd}(\alpha) = p\} \end{aligned} \tag{30.9.4}$$

where $w_O = [O^\times : \mathbb{Z}^\times]$.

Next, we claim that if $\alpha \in O$ has $\operatorname{nrd}(\alpha) = p$, then $\operatorname{trd}(\alpha) = 0$, i.e., $\alpha^2 + p = 0$. Indeed, if $t = \operatorname{trd}(\alpha)$ then the field $K = \mathbb{Q}(\alpha)$ has discriminant $t^2 - 4p < 0$ (since B is definite) so $|t| < 2\sqrt{p}$. If $t \neq 0$, then the polynomial $x^2 - tx + p$ splits modulo p, so p splits in K; but $K_p \hookrightarrow B_p$ and B_p is a division algebra, so K_p is a field, a contradiction. Thus $t = 0$.

With these results in hand, we can bring in the theory of embedding numbers. We have

$$\begin{aligned}\#\{\alpha \in O : \operatorname{nrd}(\alpha) = p\} &= \#\{\alpha \in O : \alpha^2 + p = 0\} \\ &= \sum_{S \supseteq \mathbb{Z}[\sqrt{-p}]} \#\operatorname{Emb}(S, O).\end{aligned} \tag{30.9.5}$$

The group $O^\times/\{\pm 1\}$ acts by conjugation on $\operatorname{Emb}(S, O)$ without fixed points as in 30.8.3: since $p \neq 2, 3$, we have $S^\times = \{\pm 1\}$. Thus

$$\#\operatorname{Emb}(S, O) = w_O m(S, O; O^\times). \tag{30.9.6}$$

Combining (30.9.4), (30.9.5), and (30.9.6), and plugging into (30.9.3), we have

$$\#\operatorname{Typ} O = \frac{1}{2}\#\operatorname{Cls} O + \frac{1}{4} \sum_{S \supseteq \mathbb{Z}[\sqrt{-p}]} \sum_{[O'] \in \operatorname{Typ} O} m(S, O'; O'^\times). \tag{30.9.7}$$

By (30.4.21) (rewriting Theorem 30.4.7), for $S \supseteq \mathbb{Z}[\sqrt{-p}]$ we have

$$\sum_{[O'] \in \operatorname{Typ} O} h(O') m(S, O'; O'^\times) = h(S) m(\widehat{S}, \widehat{O}; \widehat{O}^\times)$$

where $h(O') = [\operatorname{Idl} O' : \operatorname{PIdl} O']$; but $h(O') = 1$ whenever $m(S, O'; O'^\times) \neq 0$ by the first paragraph, so we may substitute into (30.9.7) to get

$$\#\operatorname{Typ} O = \frac{1}{2}\#\operatorname{Cls} O + \frac{1}{4} \sum_{S \supseteq \mathbb{Z}[\sqrt{-p}]} h(S) m(\widehat{S}, \widehat{O}; \widehat{O}^\times). \tag{30.9.8}$$

The order O is maximal, so the adelic embedding number is the product of local embedding numbers computed in Proposition 30.5.3: there is only a possible contribution at p, since $p \neq 2$ the order S is maximal, and K is ramified so $m(S_p, O_p; O_p^\times) = 1$, thus $m(\widehat{S}, \widehat{O}; \widehat{O}^\times) = 1$.

Finally, the order $\mathbb{Z}[\sqrt{-p}]$ of discriminant $-4p$ is maximal whenever $p \equiv 1 \pmod 4$ and the sum becomes simply $h(-4p)$; when $p \equiv 3 \pmod 4$, this order is contained in the maximal order of discriminant $-p$, so the sum is $h(-p) + h(-4p)$. The result is proven. □

Remark 30.9.9. For an alternate direct proof of Proposition 30.9.2 working with elliptic curves, see Cox [Cox89, Theorem 14.18].

30.9.10. The sum of class numbers in Proposition 30.9.2 can be rewritten uniformly in terms of the ring of integers as follows:

$$[h(-p)] + h(-4p) = \#\operatorname{Cl}\mathbb{Q}(\sqrt{-p}) \cdot \begin{cases} 1, & \text{if } p \equiv 1 \pmod 4; \\ 4, & \text{if } p \equiv 3 \pmod 8; \\ 2, & \text{if } p \equiv 7 \pmod 8). \end{cases}$$

When $p \equiv 1 \pmod 4$, there is nothing to do. For $p \equiv 3 \pmod 4$, we have

$$h(-4p) = \begin{cases} 3h(-p), & \text{if } p \equiv 3 \pmod 8; \\ h(-p), & \text{if } p \equiv 7 \pmod 8; \end{cases}$$

according as 2 is inert or split in $K = \mathbb{Q}(\sqrt{-p})$.

Remark 30.9.11. In section 42.1–42.2, we relate quaternion algebras to supersingular elliptic curves; in this language, Proposition 30.9.2 gives rise to a formula for the number of supersingular elliptic curves defined over $\mathbb{F}_p$ up to isomorphism.

Remark 30.9.12. Eichler [Eic56a, Satz 11] gave a type number formula for definite hereditary orders over a totally real field; this formula has an error which was corrected by Peters [Pet69, Satz 14, Satz 15] over fields of class number one and by Pizer [Piz73, Theorem A] in general. Pizer [Piz76a, Theorem 26] gives a formula for the type number for (general) Eichler orders over $\mathbb{Q}$. Finally, Vignéras [Vig80a, Corollaire V.2.6] gives a "structural" type number formula (without explicit evaluation of the sum) for Eichler orders, and Körner [Kör87, Theorem 3] gives a general type number formula. For a generalization to totally definite orders in central simple algebras of prime index over global fields, see Brzezinski [Brz97].

Exercises

1. As in section 30.1, for $k \geq 1$ let

$$r_k(n) := \#\{(x_1, x_2, \dots, x_k) \in \mathbb{Z}^k : x_1^2 + x_2^2 + \dots + x_k^2 = n\}.$$

   We gave formulas for $r_3(n), r_4(n)$. For completeness, observe that $r_1(n) = 2, 0$ according as n is a square or not, and give a formula for $r_2(n)$ in terms of the factorization of n in the ring $\mathbb{Z}[i]$.

▷ 2. Let B be a quaternion algebra over $\mathbb{Q}$, let $O \subset B$ be an order, let K be a quadratic field with an embedding $K \hookrightarrow B$ and suppose $S = K \cap O$ is the ring of integers of K.

   (a) Let $\mathfrak{b} \subset K$ be an invertible fractional S-ideal. Show that $\mathfrak{b}O \cap K = \mathfrak{b}$. *[Hint: since* $1 \in O$*, we have* $\mathfrak{b}O \cap K \supseteq \mathfrak{b}$*. For the other inclusion, consider*

$$(\mathfrak{b}O \cap K) \cdot \mathfrak{b}^{-1}\mathfrak{b} \subseteq (\mathfrak{b}\mathfrak{b}^{-1}O \cap K) \cdot \mathfrak{b} = \mathfrak{b}.]$$

   (b) Rewrite the proof in (a) idelically.

▹ 3. Let $n \equiv 1, 2 \pmod 4$.

   (a) Let p be an odd prime, and let $\alpha, \alpha' \in \mathrm{M}_2(\mathbb{Z}_p)$ satisfy $\alpha^2 + n = 0$. Show that there exists $\mu \in \mathrm{GL}_2(\mathbb{Z}_p)$ such that $\alpha' = \mu^{-1}\alpha\mu$.
   (b) Let $B_2 = \left(\dfrac{-1,-1}{\mathbb{Q}_2}\right)$ and O_2 its valuation ring. Show that if α, α' satisfy the same reduced characteristic polynomial, then there exists $\nu \in N_{B_2^\times}(O_2)$ such that $\alpha' = \nu^{-1}\alpha\nu$. *[Hint: $N_{B_2^\times}(O_2) = B_2^\times$.]*
   (c) Put together (a) and (b) to conclude (30.2.4).

4. Deduce the theorem of Gauss (Theorem 30.1.3) from Theorem 30.4.7.
5. Let $m, n \in \mathbb{Z}_{>0}$ be odd, coprime, with $m < 2\sqrt{n}$. Show that there are $12h(m^2-4n)$ solutions to the equations

$$\begin{aligned} a^2 + b^2 + c^2 + d^2 &= n \\ a + b + c + d &= m \end{aligned}$$

with $a, b, c, d \in \mathbb{Z}$. *[Hint: consider the unit $(1-i-j-k)/2$ of the Hurwitz order.]*
6. Specialize Theorem 30.1.5 to the case $D = p$ and $N = 1$, as follows. Let B be a definite quaternion algebra over $\mathbb{Q}$ of discriminant $D = p$ prime and let $O \subset B$ be a maximal order (Eichler order of level $M = 1$). Show that

$$\#\operatorname{Cls} O = \begin{cases} 1, & \text{if } p = 2, 3; \\ (p-1)/12, & \text{if } p \equiv 1 \pmod{12}; \\ (p+7)/12, & \text{if } p \equiv 5 \pmod{12}; \\ (p+5)/12, & \text{if } p \equiv 7 \pmod{12}; \text{ and} \\ (p+13)/12, & \text{if } p \equiv 11 \pmod{12}. \end{cases}$$

7. Let R be local and O be an Eichler R-order of level $\mathfrak{p}$ (so O is hereditary, but not maximal). Let K be a quadratic F-algebra and $S \subseteq K$ an R-order. Show that $m(S, O; O^\times) = \emptyset$ if and only if S is maximal and $\left(\dfrac{K}{\mathfrak{p}}\right) = -1$.
8. Let B be an indefinite quaternion algebra over $\mathbb{Q}$ of discriminant D and let $O \subseteq B$ be an Eichler order. Let $S \subseteq K$ be a quadratic order of discriminant $d_S = df^2$, where d is a fundamental discriminant.

   (a) Suppose O is maximal. Show that there exists an optimal embedding $S \hookrightarrow O$ if and only if for all $p \mid D$, we have $(K \mid p) \neq 1$ and $p \nmid f$.
   (b) Suppose O is Eichler of level M and (for simplicity) that $N := DM$ is odd. Show that there exists an optimal embedding $S \hookrightarrow O$ if and only if the following conditions hold:
   - For all $p \mid D$, we have $(K \mid p) \neq 1$.
   - For all $p^e \parallel M$, either $(K \mid p) = 1$ or $p^e \mid d_S$.

   [Hint: recall that $\#\operatorname{Typ} O = 1$.]
   [It is possible to extend this to all N by analyzing local embeddings at $p = 2$ working over $\mathbb{Q}_2$.]

9. Let $F = \mathbb{Q}(\sqrt{d})$ be a real quadratic field of discriminant $d > 0$. Show that there exists $q \neq 2, 3$ such that $[F(\zeta_{2q}) : F] = 2$ if and only if $d = 5, 8, 12$.
10. Let R be a global ring with field of fractions F, let B be a quaternion algebra over F, and let $O, O' \subseteq B$ be maximal R-orders. Let $K \subseteq B$ be a quadratic F-algebra in B, and suppose that $K \cap O = K \cap O'$; let S be this common, optimally embedded quadratic R-order. Show that there exists an invertible fractional S-ideal $\mathfrak{a} \subseteq K$ such that $O\mathfrak{a} = \mathfrak{a}O'$. *[Hint: the statement is local, so put the Eichler order $O \cap O'$ and generator $\gamma \in S$ in standard form; show that there exists $\mu_\mathfrak{p} \in O^\times \simeq \mathrm{GL}_2(R)$ and $\alpha_\mathfrak{p} \in S_\mathfrak{p}$ such that $\mu_\mathfrak{p}\alpha_\mathfrak{p} = \varpi_\mathfrak{p}$ generates the normalizer.]*
11. In this exercise, we prove a local-global principle for optimal embeddings in a self-contained manner. [For a much more precise result, read on to Chapter 31!] Let F be a number field with ring of integers R. Let B be a quaternion algebra over F and let $O \subseteq B$ be an R-order. Let K be a quadratic F-algebra and let $S \subseteq K$ be an R-order. Suppose that $K \hookrightarrow B$; equivalently, by the local-global principle for embeddings (Proposition 14.6.7), every place $v \in \mathrm{Pl}\,F$ does not split in K.

    (a) Suppose that for all primes $\mathfrak{p}$ of R, there exists an optimal embedding $\phi_\mathfrak{p} : S_\mathfrak{p} \hookrightarrow O_\mathfrak{p}$. Show that there exists an order $O' \in \mathrm{Gen}(O)$ (so O' is locally isomorphic to O) and an optimal embedding $\phi \colon S \hookrightarrow O'$. *[Hint: choose a maximal order containing S and apply finitely many local corrections.]*
    (b) Now suppose that $\#\,\mathrm{Cl}_{G(O)}\,R = 1$. Show that there exists an optimal embedding $\phi \colon S \hookrightarrow O$.
    (c) Show that parts (a) and (b) follow from the trace formula (Theorem 30.4.7).

12. Give a direct proof of Corollary 30.4.23.

This chapter is licensed under the terms of the Creative Commons Attribution-NonCommercial 4.0 International License (http://creativecommons.org/licenses/by-nc/4.0/), which permits any noncommercial use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license and indicate if changes were made.

The images or other third party material in this chapter are included in the chapter's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the chapter's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.

Chapter 31

Selectivity

In the previous chapter, we saw that (conjugacy classes of) embeddings of a quadratic order into a quaternion algebra are naturally distributed over the genus of a quaternion order; in applications, we want to compare the number of embeddings over orders in a genus. Such a comparison can be thought of as a strong integral refinement of the local-global principle for embeddings of quadratic fields (Proposition 14.6.7), which belongs to the more general framework of the Albert–Brauer–Hasse–Noether theorem.

This chapter is quite technical, and it may be skipped on a first reading. To reward the reader who persists, we conclude this chapter with the construction of isospectral, nonisometric hyperbolic Riemannian manifolds, following Vignéras.

31.1 Selective orders

To get a preview of what selectivity is all about, right off the bat we give an example of the failure for a quadratic order to embed equitably in the genus of an order.

Example 31.1.1. Let $F := \mathbb{Q}(\sqrt{-5})$ and $R := \mathbb{Z}_F = \mathbb{Z}[\sqrt{-5}]$. Then $\mathrm{Cl}\, R \simeq \mathbb{Z}/2\mathbb{Z}$, and the nontrivial class is represented by the ideal $\mathfrak{p} = \langle 2, 1 + \sqrt{-5} \rangle \subseteq \mathbb{Z}_F$ with $\mathfrak{p}^2 = 2\mathbb{Z}_F$. By class field theory, the Hilbert class field $K \supseteq F$ is a quadratic extension, and the genus theory of Gauss gives $K = F(\sqrt{-1}) = F(\sqrt{5})$. The maximal order of K is $\mathbb{Z}_K = \mathbb{Z}_F[w]$ where

$$w = \sqrt{-1}\left(\frac{1+\sqrt{5}}{2}\right) = \frac{\sqrt{-1}+\sqrt{-5}}{2}$$

satisfies $w^2 - \sqrt{-5}w - 1 = 0$, a polynomial of discriminant $-5 + 4 = -1$.

We take $B = \mathrm{M}_2(F)$ and the maximal order $O = \mathrm{M}_2(\mathbb{Z}_F)$. By 17.3.7, there is a bijection $\mathrm{Cl}\, R \xrightarrow{\sim} \mathrm{Cls}_{\mathsf{R}}\, O$, with the nontrivial right ideal class represented by

$$I = \begin{pmatrix} \mathfrak{p} & 0 \\ 0 & R \end{pmatrix} \mathrm{M}_2(R) = \begin{pmatrix} \mathfrak{p} & \mathfrak{p} \\ R & R \end{pmatrix};$$

its left order is

$$O' = O_{\mathsf{L}}(I) = \begin{pmatrix} R & \mathfrak{p} \\ \mathfrak{p}^{-1} & R \end{pmatrix}.$$

© The Author(s) 2021

J. Voight, *Quaternion Algebras*, Graduate Texts in Mathematics 288,
https://doi.org/10.1007/978-3-030-56694-4_31

These two orders are not isomorphic and up to isomorphism represent the two types of maximal R-orders in $\mathrm{M}_2(F)$.

We claim that there is an embedding $\mathbb{Z}_K \hookrightarrow O$ but *no* embedding $\mathbb{Z}_K \hookrightarrow O'$. The first part of the claim is easy: taking the rational canonical form, we take the embedding

$$w \mapsto \alpha = \begin{pmatrix} 0 & 1 \\ 1 & \sqrt{-5} \end{pmatrix}. \tag{31.1.2}$$

The proof that $\mathbb{Z}_K \not\hookrightarrow O'$ is more difficult. (The embedding (31.1.2) does not extend to O' because of the off-diagonal coefficients; and we cannot conjugate this embedding in an obvious way because the ideal $\mathfrak{p}$ is not principal.) Such an embedding would be specified by a matrix

$$\alpha' = \begin{pmatrix} a & b \\ c & -a+\sqrt{-5} \end{pmatrix} \in \begin{pmatrix} R & \mathfrak{p} \\ \mathfrak{p}^{-1} & R \end{pmatrix}$$

with

$$-\det(\alpha') = a^2 - \sqrt{-5}a + bc = 1; \tag{31.1.3}$$

so the content in the second claim is that there is *no* solution to the quadratic equation (31.1.3).

Indeed, suppose there is a solution. Let $f(x) = x^2 - \sqrt{-5}x - 1 \in \mathbb{Z}_F[x]$, so that $f(a) + bc = 0$. We may factor $b\mathbb{Z}_F = \mathfrak{p}\mathfrak{b}$ with $\mathfrak{b} \subseteq \mathbb{Z}_F$ and $[\mathfrak{b}] \in \mathrm{Cl}\,\mathbb{Z}_F$ nontrivial; by parity, there exists a prime $\mathfrak{q} \mid \mathfrak{b}$ with $[\mathfrak{q}]$ nontrivial. Factoring $c\mathbb{Z}_F = \mathfrak{p}^{-1}\mathfrak{c}$ with $\mathfrak{c} \subseteq \mathbb{Z}_F$, we have $bc\mathbb{Z}_F = \mathfrak{b}\mathfrak{c} \subseteq \mathfrak{q}$, so $f(a) = -bc \equiv 0 \pmod{\mathfrak{q}}$. But $f(x)$ has trivial discriminant, and modulo a prime $\mathfrak{q} \subseteq \mathbb{Z}_F$ it either splits (into distinct linear factors) or remains irreducible. And by the Artin map, $f(x)$ splits modulo $\mathfrak{q}$ if and only if $\mathfrak{q}$ splits in K if and only if the class $[\mathfrak{q}] \in \mathrm{Cl}\,\mathbb{Z}_F$ is trivial. Putting these two pieces together, we have $f(a) \equiv 0 \pmod{\mathfrak{q}}$ and $f(x)$ is irreducible modulo $\mathfrak{q}$. This is a contradiction, and there can be no solution.

With this cautionary but illustrative example in hand, we state our main theorem. We return to the idelic notation of section 30.4. We will consider embeddings in the context of strong approximation (see Chapter 28).

The following notation will be in use throughout this chapter.

31.1.4. Let $R = R_{(T)}$ be a global ring with eligible set T and let $F = \mathrm{Frac}\,R$ be its field of fractions. Let B be a quaternion algebra over F and suppose that B is T-indefinite. Let $O \subseteq B$ be an R-order.

Let $K \supseteq F$ be a separable quadratic F-algebra and let $S \subseteq K$ be an R-order. Suppose that $\mathrm{Emb}(\widehat{S}; \widehat{O}) \neq \emptyset$, which is to say, for all primes $\mathfrak{p} \subseteq R$, the $R_\mathfrak{p}$-algebra $S_\mathfrak{p}$ embeds *optimally* into $O_\mathfrak{p}$.

Our struggle will be to understand when the local optimal embeddings glue together to give a global optimal embedding. As a start, we know by Corollary 30.4.18 that there exists some order $O' \in \mathrm{Gen}\,O$ in the genus of O (i.e., locally isomorphic to O) such that $\mathrm{Emb}(S, O') \neq \emptyset$.

Definition 31.1.5. We say that $\operatorname{Gen} O$ is **genial** for S if $\operatorname{Emb}(S, O') \neq \emptyset$ for all $O' \in \operatorname{Gen} O$. If $\operatorname{Gen} O$ is not genial, i.e., there exists $O' \in \operatorname{Gen} O$ such that

$$\operatorname{Emb}(S, O') = \emptyset,$$

then we say that $\operatorname{Gen} O$ is **optimally selective** for S.

By definition, $\operatorname{Gen} O$ is genial for S if and only if S embeds optimally in every order O' that is locally isomorphic to O.

31.1.6. We define the following condition, called the **optimal selectivity condition**:

(OS) K is a subfield of the class field $H_{GN(O)}$ of F obtained from $\operatorname{Cl}_{GN(O)} R$.

In particular, if K is not a field, then (OS) does not hold. We now state our main theorem, with notation and hypotheses in 31.1.4.

Main Theorem 31.1.7 (Optimal selectivity)**.** *Suppose that O is an Eichler order. Then the following statements hold.*

(a) $\operatorname{Gen} O$ *is optimally selective for S if and only if the optimal selectivity condition* (OS) *holds.*
(b) *If* $\operatorname{Gen} O$ *is optimally selective for S, then* $\operatorname{Emb}(S, O') \neq \emptyset$ *for precisely* half *of the types* $[O'] \in \operatorname{Typ} O$.
(c) *In all cases,*

$$m(S, O'; O'^{\times}) = m(S, O; O^{\times})$$

for all $O' \in \operatorname{Gen} O$ *whenever both sides are nonzero.*

Since the optimal selectivity condition (OS) only depends on K, if $\operatorname{Gen} O$ is optimally selective for S then it is optimally selective for all R-orders in K.

Remark 31.1.8. It was first noted by Chevalley [Chev36] in the more general situation of matrix algebras that it was possible for a commutative order to embed into some, but not all, maximal orders. An approach to selectivity is sketched by Vignéras [Vig80a, Théorème III.5.15], but there are some glitches [CF99, Remark 3.4]. Maclachlan [Macl2008, Theorem 1.4] gives a proof of Main Theorem 31.1.7(a)–(b) for hereditary orders (Eichler orders of squarefree level). For a more detailed literature survey and further comments, see 31.7.7.

31.1.9. When $\operatorname{Gen} O$ *is* optimally selective for S, then we can refine Main Theorem 31.1.7(b) detecting the half of types of orders for which there is an optimal embedding of S.

From our hypothesis $\operatorname{Emb}(\widehat{S}, \widehat{O}) \neq \emptyset$, we know that S embeds into some order in the genus of O; we might as well take this to be O itself, so we suppose that $\operatorname{Emb}(S, O) \neq \emptyset$. Let $O' \in \operatorname{Gen} O$. Then O' is connected to O, so $O' = O_{\mathsf{L}}(I)$ for an invertible right O-ideal $I \subseteq O$. Let

$$\mathfrak{a} = [O : O \cap O'] = [O' : O \cap O'].$$

Then $\operatorname{Emb}(S, O') \neq \emptyset$ if and only if $\operatorname{Frob}_{\mathfrak{a}}$ is trivial in $\operatorname{Gal}(K \,|\, F)$.

For example, if $\mathfrak{a} = \mathfrak{p}$ is prime, then $\operatorname{Frob}_{\mathfrak{p}}$ is trivial in $\operatorname{Gal}(K \,|\, F)$ if and only if $\mathfrak{p}$ is not inert in K. For maximal orders, we can equivalently formulate the index in terms of distance on the Bruhat–Tits tree (see section 23.5 and Exercise 23.9).

The core application of the optimal selectivity theorem is the following corollary.

Corollary 31.1.10. *Suppose that* Gen O *is genial for* S. *Then*

$$m(S, O; O^\times) = \frac{h(S)}{\#\operatorname{Cls} O} m(\widehat{S}, \widehat{O}; \widehat{O}^\times).$$

Proof. We combine Main Theorem 31.1.7 and Theorem 30.4.7. □

We conclude this introduction with a second application, a generalization of Corollary 28.6.4. Let Ω be the set of real, ramified places of B, and let $R^\times_{>_\Omega 0} := R^\times \cap F^\times_{>_\Omega 0}$ be the subgroup of $R^\times$ of elements that are positive at all $v \in \Omega$.

Corollary 31.1.11. *Let* $O \subseteq B$ *be an Eichler* R*-order. Then*

$$\operatorname{nrd}(O^\times) = R^\times_{>_\Omega 0}.$$

Proof. Let $u \in R^\times_{>_\Omega 0}$. We repeat the argument of Corollary 28.6.4: we find $\gamma' \in O'$ with $\operatorname{nrd}(\gamma') = u$ and $O' \in \operatorname{Gen} O$. We may suppose further that Gen O is not selective for $R[\gamma']$ by shrinking the open set to ensure that $K = \operatorname{Frac} R[\gamma'] \not\subseteq H_{GN(O)}$. Let $S = K \cap O'$. Then $S \subseteq O'$ is optimally embedded; and by Main Theorem 31.1.7(c), there exists an optimal embedding $\phi : S \hookrightarrow O$, hence $\phi(\gamma') = \gamma \in O$ has $\operatorname{nrd}(\gamma) = u$ as desired. □

31.2 Selectivity conditions

In this brief section, we make the somewhat opaque optimal selectivity condition (OS) explicit for Eichler orders.

Proposition 31.2.1. *Let* O *be an Eichler order of level* $\mathfrak{M}$. *Then Condition* (OS) *holds if and only if all of the following four conditions hold:*

(a) *The extension* $K \supseteq F$ *and the quaternion algebra* B *are ramified at the same (possibly empty) set of archimedean places of* F*;*
(b) K *and* B *are unramified at all nonarchimedean places* $v \in \operatorname{Pl} F$*;*
(c) *Every nonarchimedean place* $v \in T$ *splits in* K*; and*
(d) *If* $\mathfrak{p} \subset R$ *is a nonzero prime and* $\operatorname{ord}_\mathfrak{p}(\mathfrak{M})$ *is odd, then* $\mathfrak{p}$ *splits in* K.

Proof. We determine the class field $H_{GN(O)}$ obtained from the group $GN(O) = F^\times_{>_\Omega 0} \operatorname{nrd}(N_{\widehat{B}^\times}(\widehat{O}))$.

Recall we have $G(O) = F^\times_{>_\Omega 0} \operatorname{nrd}(\widehat{O}^\times) = F^\times_{>_\Omega 0} \widehat{R}^\times$, since O is an Eichler order and therefore locally norm-maximal, so $H_{G(O)}$ is the maximal abelian extension of F unramified away from the real places in $\operatorname{Ram}(B)$ and such that the remaining places $v \in T$ split completely.

The normalizer $\operatorname{nrd}(N_{\widehat{B}^\times}(\widehat{O}))$ is the restricted direct product of local normalizers, computed in (23.2.8) for $\mathfrak{p} \mid \mathfrak{D}$ and Corollary 23.3.14 for $\mathfrak{p} \nmid \mathfrak{D}$: for the latter,

$$\operatorname{nrd}(N_{B^\times_\mathfrak{p}}(O_\mathfrak{p})) = \begin{cases} F^{\times 2}_\mathfrak{p} R^\times_\mathfrak{p}, & \text{if } \operatorname{ord}_\mathfrak{p}(\mathfrak{M}) \text{ is even;} \\ F^\times_\mathfrak{p}, & \text{if } \operatorname{ord}_\mathfrak{p}(\mathfrak{M}) \text{ is odd.} \end{cases}$$

Therefore, the quotient $\mathrm{Cl}_{G(O)} R \to \mathrm{Cl}_{GN(O)} R$ factors through the quotient by squares $\mathrm{Cl}_{G(O)} R/(\mathrm{Cl}_{G(O)} R)^2$ and then the further quotient by the primes $\mathfrak{p} \mid \mathfrak{D} = \operatorname{disc} B$ and $\mathfrak{p} \mid \mathfrak{M}$ with $\operatorname{ord}_{\mathfrak{p}}(\mathfrak{M})$ *odd*.

We now conclude the proof. If K is not a field then all of the desired conclusions hold; so suppose K is a field. Since $K \hookrightarrow B$, if $v \in \mathrm{Pl}\, F$ ramifies in B then v also ramifies in B. A containment $K \subseteq H_{GN(O)}$ is permitted at archimedean places if and only if the archimedean ramification in $K \supseteq F$ is no bigger than this. In a similar way, the conditions in the previous paragraph establish (c)–(d), and K is unramified at all nonarchimedean places v. To conclude (b), if $\mathfrak{p} \mid \mathfrak{D}$ then $\mathfrak{p}$ splits in $H_{GN(O)}$ and therefore in K; but we are assuming that $K \hookrightarrow B$ so $K_{\mathfrak{p}} \hookrightarrow B_{\mathfrak{p}}$, a contradiction since $B_{\mathfrak{p}}$ is a division algebra, and there can be no such $\mathfrak{p}$. □

31.3 ∗ Selectivity setup

We now embark on a proof of the selectivity theorem (Main Theorem 31.1.7); this goal will occupy us for the remainder of this chapter. In this section, we begin to isolate the problem: there is a group that is at worst $\mathbb{Z}/2\mathbb{Z}$ and is usually trivial, and we pin it down using strong approximation, the reduced norm, and class field theory. Our basic reference is Vignéras [Vig80a, Théorème III.5.15], and the surrounding text.

Our notation is as in 31.1.4.

31.3.1. To establish the main theorem in the case where $K \simeq F \times F$ is straightforward. We leave this case as an exercise (Exercise 31.1).

We suppose throughout the rest of this chapter that K is a field.

Let $O^1 \le \Gamma \le N_{B^\times}(O)$ (as in 30.3.10). Recall that there is a bijection (30.3.13)

$$\mathrm{Emb}(S, O; \Gamma) \leftrightarrow K^\times \backslash E/\Gamma$$

where

$$E = \{\beta \in B^\times : K^\beta \cap O = S^\beta\}$$

and we abbreviate conjugation $K^\beta = \beta^{-1} K \beta$ for conciseness. Conjugating if necessary, we may suppose that $1 \in E$, i.e., we start with an order and an optimal embedding $K \cap O = S$.

We employ idelic notation as in section 30.4. The inclusion $B^\times \hookrightarrow \widehat{B}^\times$ gives an inclusion

$$E/\Gamma \hookrightarrow \widehat{E}/\widehat{\Gamma} \tag{31.3.2}$$

with

$$\widehat{E} := \left\{\widehat{\beta} \in \widehat{B}^\times : \widehat{K}^{\widehat{\beta}} \cap \widehat{O} = \widehat{S}^{\widehat{\beta}}\right\}.$$

The hypothesis of strong approximation allows us to identify precisely the image of the map (31.3.2) via the reduced norm in the following way.

31.3.3. As in Theorem 28.5.5 (a motivating application of strong approximation), the reduced norm induces a bijection

$$B^\times \backslash \widehat{B}^\times / \widehat{\Gamma} \xrightarrow{\sim} F^\times_{>_\Omega 0} \backslash \widehat{F}^\times / \operatorname{nrd}(\widehat{\Gamma}) = \mathrm{Cl}_{G(\Gamma)} R \tag{31.3.4}$$

where $G(\Gamma) := F^\times_{>_\Omega 0} \operatorname{nrd}(\widehat{\Gamma})$.

Lemma 31.3.5. *We have*

$$E/\Gamma = \{\widehat{\beta\Gamma} \in \widehat{E}/\widehat{\Gamma} : \mathrm{nrd}(\widehat{\beta}) \in G(\Gamma)\} \subseteq \widehat{E}/\widehat{\Gamma}.$$

That is to say, if $\widehat{\beta} \in \widehat{E}$*, then there exists* $\beta \in E$ *such that* $\widehat{\beta\Gamma} = \beta\widehat{\Gamma}$ *if and only if* $\mathrm{nrd}(\widehat{\beta}) \in G(\Gamma)$.

Proof. We find a $\beta \in B^\times$ (without the condition that $\beta \in E$) immediately from the bijection (31.3.4). But $\beta = \widehat{\beta}\widehat{\gamma} \in \widehat{\beta\Gamma}$ and $\widehat{\Gamma} \leq N_{\widehat{B}^\times}(\widehat{O})$ gives

$$\widehat{K}^\beta \cap \widehat{O}^{\widehat{\gamma}} = \widehat{K}^\beta \cap \widehat{O} = \widehat{S}^\beta$$

and intersecting with B we find $\beta \in E$. □

Lemma 31.3.5 points the way more generally, at least to detect if there is an embedding in the first place in an order. First, we need to give representatives of the type set.

31.3.6. Recalling 28.5.7, there is a bijection

$$\mathrm{Typ}\, O \leftrightarrow B^\times \backslash \widehat{B}^\times / N_{\widehat{B}^\times}(\widehat{O});$$

explicitly, every isomorphism class of order in $\mathrm{Typ}\, O$ is of the form

$$O' = \widehat{\nu}\widehat{O}\widehat{\nu}^{-1} \cap B = \widehat{O}^{\widehat{\nu}^{-1}} \cap B$$

(yes, the choice of inverse is deliberate), with the class of $\widehat{\nu} \in \widehat{B}^\times$ in $B^\times \backslash \widehat{B}^\times / N_{\widehat{B}^\times}(\widehat{O})$ uniquely defined.

In the presence of strong approximation (Corollary 28.5.10), we have a further bijection

$$\mathrm{Typ}\, O \leftrightarrow \mathrm{Cl}_{GN(O)}\, R$$

where

$$GN(O) = G(N_{B^\times}(O)) = F^\times_{>_\Omega 0}\, \mathrm{nrd}(N_{\widehat{B}^\times}(\widehat{O})).$$

Now we look back at embeddings and reduced norms.

31.3.7. Let

$$\mathrm{nrd}(\widehat{E}) = \{\mathrm{nrd}(\widehat{\beta}) : \widehat{\beta} \in \widehat{E}\} \subseteq \widehat{F}^\times. \tag{31.3.8}$$

The set $\widehat{E}$ does not obviously have a group structure, so $\mathrm{nrd}(\widehat{E}) = \{\mathrm{nrd}(\widehat{\beta}) : \widehat{\beta} \in \widehat{E}\}$ is a subset. But $\widehat{E}$ is stable under $\widehat{K}^\times$, so $\mathrm{nrd}(\widehat{E})$ is a *union of cosets* of $\mathrm{nrd}(\widehat{K}^\times)$.

The set $\mathrm{nrd}(\widehat{E})$ is quite large, because it contains reduced norms from $\widehat{K}^\times \subseteq \widehat{E}$ and $N_{\widehat{B}^\times}(\widehat{O}) \subseteq \widehat{E}$.

31.3.9. By the main theorem of class field theory (Theorem 27.5.10), the Artin map gives a bijection

$$\underline{F}^\times / F^\times \mathrm{Nm}_{K|F}(\underline{K}^\times) \xrightarrow{\sim} \mathrm{Gal}(K\,|\,F) \simeq \mathbb{Z}/2\mathbb{Z}. \tag{31.3.10}$$

Let $S \subseteq T$ be the set of archimedean places of F. Recall the isomorphism (27.5.8)

$$\underline{F}^\times / F^\times F^\times_{\infty,>0} \xrightarrow{\sim} \underline{F}^\times_{S\!\!\!/} / F^\times_{>0}.$$

We have

$$F^\times_{\infty,>0} \leq \mathrm{Nm}_{K|F}(K^\times_\infty) \leq \mathrm{Nm}_{K|F}(\underline{K}^\times)$$

(the latter embedded at the infinite place with the other components 1), and by the same argument as in Lemma 27.5.6, the image of $F^\times \mathrm{Nm}_{K|F}(\underline{K}^\times)$ under the isomorphism (27.5.8) is $F^\times_{>_\Sigma 0} \mathrm{Nm}_{K|F}(K^\times_{S\!\!\!/})$ where $\Sigma \subseteq \mathrm{Pl}(F)$ is the set of places ramified in K (going from real to complex in the extension $F \subseteq K$) and

$$F^\times_{>_\Sigma 0} := \{a \in F^\times : v(a) > 0 \text{ for all real } v \in \Sigma\}.$$

Therefore we have an isomorphism

$$\underline{F}^\times / F^\times \mathrm{Nm}_{K|F}(\underline{K}^\times) \xrightarrow{\sim} \underline{F}^\times_{S\!\!\!/} / F^\times_{>_\Sigma 0} \mathrm{Nm}_{K|F}(K^\times_{S\!\!\!/}). \tag{31.3.11}$$

We then further project from the target of (31.3.11) to $\widehat{F}^\times = \underline{F}^\times_{T\!\!\!/}$ to obtain the map

$$\underline{F}^\times / F^\times \mathrm{Nm}_{K|F}(\underline{K}^\times) \to \widehat{F}^\times / F^\times_{>_\Sigma 0} \mathrm{Nm}_{K|F}(\widehat{K}^\times). \tag{31.3.12}$$

Lemma 31.3.13. *We have*

$$F^\times_{>_\Sigma 0} \mathrm{Nm}_{K|F}(\widehat{K}^\times) \leq \widehat{F}^\times$$

with total index at most 2*, and index equal to* 2 *if and only if every nonarchimedean place* $v \in T$ *is split in* K.

Proof. In the projection (31.3.12), we start with a group of order 2; in order to keep it this size, the projection away from the nonarchimedean places in T must be an isomorphism, which holds if and only if for all nonarchimedean places $v \in T$ we must have v split in K. ☐

We conclude this setup section with an overview.

31.3.14. Selectivity arises from an examination of layers in the following *selectivity sandwich*:

$$F^\times_{>_\Sigma 0} \mathrm{nrd}(\widehat{K}^\times) \overset{\text{(OS)}}{\leq} F^\times_{>_\Omega 0} \mathrm{nrd}(\widehat{K}^\times) \mathrm{nrd}(N_{\widehat{B}^\times}(\widehat{O})) \overset{m}{\leq} F^\times_{>_\Omega 0} \mathrm{nrd}(\widehat{E}) \overset{s}{\leq} \widehat{F}^\times \tag{31.3.15}$$

In terms of the sandwich bread, the left-most group $F^\times_{>_\Sigma 0} \mathrm{nrd}(\widehat{K}^\times)$ has index at most 2 in the right-most group $\widehat{F}^\times$ by Lemma 31.3.13. The set $F^\times_{>_\Omega 0} \mathrm{nrd}(\widehat{E})$ is stable under multiplication by $F^\times_{>_\Sigma 0} \mathrm{nrd}(\widehat{K}^\times)$ so is a union of its cosets in $\widehat{F}^\times$; therefore, it is actually a *subgroup*.

Much ado about a (possible) group of order two!

Again by Lemma 31.3.13, the sandwich collapses if there is a nonarchimedean place $v \in T$ that is inert in $K \supseteq F$, so there is only work to do when every $v \in T$ is

split. Under this assumption, we will show in Lemma 31.4.1 that the first inequality labelled (OS) is an equality if and only if the optimal selectivity condition (OS) holds. In Propositions 31.5.1 and 31.5.7, we will show that the middle inequality labelled m is *always* an equality and that such an equality implies equality of embedding numbers (when they are nonzero). Last but not least, in Proposition 31.4.4 we will show that the final inequality labelled s is an equality if and only if there is no selectivity obstruction, i.e., $\mathrm{Emb}(S, O') \neq \emptyset$ for all $O' \in \mathrm{Gen}\, O$.

31.4 $*$ Outer selectivity inequalities

In this section, we consider the outer ends of the selectivity sandwich 31.3.14.

The left-most inequality is interpreted in the language of class field theory as follows.

Lemma 31.4.1. *We have*

$$F^\times_{>_\Sigma 0}\,\mathrm{nrd}(\widehat{K}^\times) \leq F^\times_{>_\Omega 0}\,\mathrm{nrd}(\widehat{K}^\times)\,\mathrm{nrd}(N_{\widehat{B}^\times}(\widehat{O})) \tag{31.4.2}$$

with index at most 2, *and equality holds if and only if either the optimal selectivity condition* (OS) *holds or there exists a nonarchimedean place $v \in T$ inert or ramified in K.*

Proof. Recall that $GN(O) = F^\times_{>_\Omega 0}\,\mathrm{nrd}(N_{\widehat{B}^\times}(\widehat{O}))$. In the side sandwich

$$F^\times_{>_\Sigma 0}\,\mathrm{nrd}(\widehat{K}^\times) \leq F^\times_{>_\Omega 0}\,\mathrm{nrd}(\widehat{K}^\times)\,\mathrm{nrd}(N_{\widehat{B}^\times}(\widehat{O})) \leq \widehat{F}^\times \tag{31.4.3}$$

we again have total index at most 2. By class field theory and the Galois correspondence relative to the corresponding tower of class groups, we have

$$F^\times_{>_\Sigma 0}\,\mathrm{nrd}(\widehat{K}^\times) = F^\times_{>_\Omega 0}\,\mathrm{nrd}(\widehat{K}^\times)\,\mathrm{nrd}(N_{\widehat{B}^\times}(\widehat{O})) < \widehat{F}^\times$$

(so first equality, then strict inequality in (31.4.3)) if and only if $K \subseteq H_{GN(O)}$ if and only if (OS) holds. The result then follows by Lemma 31.3.13. ∎

We next consider the right-most inequality, and we show that it contains the obstruction to selectivity.

Proposition 31.4.4. *Let* $[O'] \in \mathrm{Typ}\, O$ *be represented by the class* $B^\times \widehat{\nu} N_{\widehat{B}^\times}(\widehat{O})$. *Then* $\mathrm{Emb}(S, O') \neq \emptyset$ *if and only if* $\mathrm{nrd}(\widehat{\nu}) \in F^\times_{>_\Omega 0}\,\mathrm{nrd}(\widehat{E})$.

Proof. Define

$$E' = \{\beta' \in B^\times : K^{\beta'} \cap O' = S^{\beta'}\}$$

so that $\mathrm{Emb}_R(S, O') \xrightarrow{\sim} K^\times \backslash E'$, and similarly $\widehat{E}'$.

Suppose $\mathrm{Emb}(S, O') \neq \emptyset$, represented by $\beta' \in E'$. Then

$$\widehat{K}^{\beta'} \cap \widehat{O}' = \widehat{K}^{\beta'} \cap \widehat{O}^{\widehat{\nu}^{-1}} = \widehat{S}^{\beta'}$$

so $\beta'\widehat{\nu} = \widehat{\beta} \in \widehat{E}$. Therefore

$$\operatorname{nrd}(\widehat{\nu}) = \operatorname{nrd}(\beta'^{-1})\operatorname{nrd}(\widehat{\beta}) \in F^\times_{>_\Omega 0}\operatorname{nrd}(\widehat{E}).$$

Conversely, suppose that $\operatorname{nrd}(\widehat{\nu}) \in F^\times_{>_\Omega 0}\operatorname{nrd}(\widehat{E})$; then there exists $a \in F^\times_{>_\Omega 0}$ and $\widehat{\beta} \in \widehat{E}$ such that $\operatorname{nrd}(\widehat{\nu}) = a\operatorname{nrd}(\widehat{\beta})$. Since $\widehat{\beta} \in \widehat{E}$, we have

$$\widehat{K}^{\widehat{\beta}} \cap \widehat{O} = \widehat{S}^{\widehat{\beta}}$$

thus if $\widehat{\beta}' = \widehat{\beta}\widehat{\nu}^{-1}$ we get

$$\widehat{K}^{\widehat{\beta}'} \cap \widehat{O}^{\widehat{\nu}^{-1}} = \widehat{K}^{\widehat{\beta}'} \cap \widehat{O}' = \widehat{S}^{\widehat{\beta}'} \tag{31.4.5}$$

and $\widehat{\beta}' \in \widehat{E}'$. We have

$$\operatorname{nrd}(\widehat{\beta}') = \operatorname{nrd}(\widehat{\beta}\widehat{\nu}^{-1}) = a^{-1} \in F^\times_{>_\Omega 0}$$

So by Lemma 31.3.5, there exists $\beta' \in E'$ mapping to $\widehat{\beta}'$, and $\operatorname{Emb}(S, O') \neq \emptyset$ as claimed. □

The following corollary indicates the significance of the preceding proposition.

Corollary 31.4.6. *If* $F^\times_{>_\Omega 0}\operatorname{nrd}(\widehat{E}) = \widehat{F}^\times$, *then* $\operatorname{Emb}(S, O') \neq \emptyset$ *for all orders* $O' \in \operatorname{Gen} O$ *in the genus of* O. *Otherwise,* $F^\times_{>_\Omega 0}\operatorname{nrd}(\widehat{E}) < \widehat{F}^\times$ *has index* 2 *and* $\operatorname{Emb}(S, O') \neq \emptyset$ *for precisely* half *of the types of orders in* $\operatorname{Typ} O$*: we have* $\operatorname{Emb}(S, O') \neq \emptyset$ *for*

$$O' = \widehat{\nu}O\widehat{\nu}^{-1} \cap B$$

if and only if $\operatorname{nrd}(\widehat{\nu}) \in F^\times_{>_\Omega 0}\operatorname{nrd}(\widehat{K}^\times)$.

In particular, in the latter case we have $\#\operatorname{Typ} O$ even.

Proof. We apply Proposition 31.4.4, with indexing of the type set as in 31.3.6. □

31.5 ∗ Middle selectivity equality

In this section, we pursue the middle inequality in the selectivity sandwich 31.3.14.

First, we show that equality in this middle equality implies equality of embedding numbers, whenever they are nonzero.

Proposition 31.5.1. *We have*

$$F^\times_{>_\Omega 0}\operatorname{nrd}(\widehat{K}^\times)\operatorname{nrd}(N_{\widehat{B}^\times}(\widehat{O})) \leq F^\times_{>_\Omega 0}\operatorname{nrd}(\widehat{E}) \tag{31.5.2}$$

with index at most 2. *If equality holds in* (31.5.2)*, then whenever* $O' \in \operatorname{Gen} O$ *and* $\operatorname{Emb}(S, O')$ *is* nonempty*, we have*

$$m(S, O; O^\times) = m(S, O'; O'^\times).$$

Proof. The statement about index follows from the layering of the sandwich (31.3.15). For the second statement, suppose that $\operatorname{Emb}(S, O')$ is nonempty; then by Proposition 31.4.4, we have $O' = \widehat{O}^{\widehat{\nu}^{-1}} \cap B$ with $\operatorname{nrd}(\widehat{\nu}) \in F^{\times}_{>_\Omega 0} \operatorname{nrd}(\widehat{E})$. If equality holds in (31.5.2), then there exists $a \in F^{\times}_{>_\Omega 0}$, $\widehat{\alpha} \in \widehat{K}^{\times}$, and $\widehat{\eta} \in N_{\widehat{B}^\times}(\widehat{O})$ such that

$$\operatorname{nrd}(\widehat{\nu}) = a \operatorname{nrd}(\widehat{\alpha}) \operatorname{nrd}(\widehat{\eta}). \tag{31.5.3}$$

We restore notation from Proposition 31.4.4, and modify the argument in the converse. We define the map

$$\begin{aligned} E &\to \widehat{E}' \\ \beta &\mapsto \widehat{\beta}' = \widehat{\alpha}\beta\widehat{\eta}\widehat{\nu}^{-1} \end{aligned} \tag{31.5.4}$$

We argue as in (31.4.5). From $\beta \in E$, we have $\widehat{K}^\beta \cap \widehat{O} = \widehat{S}^\beta$. We have $\widehat{\alpha} \in \widehat{K}^\times$, so $\widehat{K}^{\widehat{\alpha}} = \widehat{K}$. And $\widehat{\eta} \in N_{\widehat{B}^\times}(\widehat{O})$, so $\widehat{O}^{\widehat{\eta}} = \widehat{O}$. Therefore

$$\widehat{K}^{\widehat{\beta}'} \cap \widehat{O}^{\widehat{\nu}^{-1}} = \widehat{K}^{\widehat{\beta}'} \cap \widehat{O}' = \widehat{S}^{\widehat{\beta}'} \tag{31.5.5}$$

and indeed $\widehat{\beta}' \in \widehat{E}'$. Finally, by (31.5.3)

$$\operatorname{nrd}(\widehat{\beta}') = \operatorname{nrd}(\widehat{\alpha}\beta\widehat{\eta}\widehat{\nu}^{-1}) = a^{-1} \operatorname{nrd}(\beta) \in F^{\times}_{>_\Omega 0}.$$

By Lemma 31.3.5, there exists $\beta' \in E'$ such that $\beta'\widehat{O}'^\times = \widehat{\beta}'\widehat{O}'^\times$, well-defined up to $O^\times$. Therefore, (31.5.4) descends to a map $E \to E'/O'^\times$, and it further descends to a map

$$\begin{aligned} E/O^\times &\to E'/O'^\times \\ \beta O^\times &\mapsto \beta' O'^\times \end{aligned} \tag{31.5.6}$$

because

$$\widehat{\alpha}\beta\mu\widehat{\eta}\widehat{\nu}^{-1}\widehat{O}' = \widehat{\alpha}\beta\mu\widehat{\eta}\widehat{O}\widehat{\nu}^{-1} = \widehat{\alpha}\beta\mu\widehat{O}\widehat{\eta}\widehat{\nu}^{-1} = \widehat{\alpha}\beta\widehat{O}\widehat{\eta}\widehat{\nu}^{-1} = \widehat{\alpha}\beta\widehat{\eta}\widehat{\nu}^{-1}\widehat{O}'.$$

This map works as well interchanging the roles of O and O', and after a little chase, we verify that the map (31.5.6) is bijective. Taking orbits under $K^\times$ on the left, we conclude the proof. ∎

In fact, equality holds in the middle for Eichler orders.

Proposition 31.5.7. *If O is an Eichler order, then the inequality* (31.5.2) *is an equality.*

Proof. We will prove that

$$\operatorname{nrd}(\widehat{K}^\times) \operatorname{nrd}(N_{\widehat{B}^\times}(\widehat{O})) = \operatorname{nrd}(\widehat{E}); \tag{31.5.8}$$

the inclusion ($\leq$) was direct, so we prove ($\geq$). The desired equality is now idelic, so we reduce to checking in the completion at a prime $\mathfrak{p}$. If $B_\mathfrak{p}$ is a division algebra, then $O_\mathfrak{p}$ is maximal, and $m(S_\mathfrak{p}, O_\mathfrak{p}; N_{B_\mathfrak{p}^\times}(O_\mathfrak{p})) = 1$ by Proposition 30.5.3(b) we have the stronger equality $E_\mathfrak{p} = N_{B_\mathfrak{p}^\times}(O_\mathfrak{p}) = B_\mathfrak{p}^\times$.

Otherwise, $B_{\mathfrak{p}} \simeq \mathrm{M}_2(F_{\mathfrak{p}})$ is split. Without loss of generality, we may suppose that $O_{\mathfrak{p}}$ is a standard Eichler order. Further, by Lemma 30.6.3, after conjugating by a normalizing element if necessary, we may suppose that the reference optimal embedding $S_{\mathfrak{p}} \hookrightarrow O_{\mathfrak{p}}$ is normalized. But then this embedding extends to an optimal embedding $S_{\mathfrak{p}} \hookrightarrow \mathrm{M}_2(R_{\mathfrak{p}})$: the upper-right entry is 1.

Now let $\beta_{\mathfrak{p}} \in E_{\mathfrak{p}}$ be arbitrary, with associated embedding $\phi_{\mathfrak{p}}^{\beta_{\mathfrak{p}}} : S_{\mathfrak{p}} \hookrightarrow O_{\mathfrak{p}}$. We repeat the argument in the previous paragraph: by Lemma 30.6.3, replacing $\beta_{\mathfrak{p}}$ by $\beta_{\mathfrak{p}}\nu_{\mathfrak{p}}$ if necessary with $\nu_{\mathfrak{p}} \in N_{B_{\mathfrak{p}}^\times}(O_{\mathfrak{p}})$, we may suppose that $\phi_{\mathfrak{p}}^{\beta_{\mathfrak{p}}}$ is normalized, and therefore extends to an optimal embedding into $\mathrm{M}_2(R_{\mathfrak{p}})$. But by Proposition 30.5.3(a), we have $m(S_{\mathfrak{p}}, \mathrm{M}_2(R_{\mathfrak{p}}); \mathrm{GL}_2(R_{\mathfrak{p}})) = 1$—all optimal embeddings into $\mathrm{M}_2(R_{\mathfrak{p}})$ are conjugate under $\mathrm{GL}_2(R_{\mathfrak{p}})$—so there exists $\mu_{\mathfrak{p}} \in \mathrm{GL}_2(R_{\mathfrak{p}})$ such that $\phi_{\mathfrak{p}}^{\beta_{\mathfrak{p}}} = \phi_{\mathfrak{p}}^{\mu_{\mathfrak{p}}}$. Therefore $\beta_{\mathfrak{p}} \in \mu_{\mathfrak{p}} K_{\mathfrak{p}}^\times$ so

$$\mathrm{nrd}(\beta_{\mathfrak{p}}) \in \mathrm{nrd}(K_{\mathfrak{p}}^\times) R_{\mathfrak{p}}^\times \le \mathrm{nrd}(K_{\mathfrak{p}}^\times)\,\mathrm{nrd}(O_{\mathfrak{p}}^\times) \le \mathrm{nrd}(K_{\mathfrak{p}}^\times)\,\mathrm{nrd}(N_{B_{\mathfrak{p}}^\times}(O_{\mathfrak{p}})) \tag{31.5.9}$$

as claimed. □

31.6 ∗ Optimal selectivity conclusion

We now officially complete the proof of the selectivity theorem for Eichler orders.

Proof of Main Theorem 31.1.7. By 31.3.1, we may suppose K is a field. We refer to the selectivity sandwich (31.3.15), using Proposition 31.5.7 to simplify the middle equality:

$$F_{>_\Sigma 0}^\times \,\mathrm{nrd}(\widehat{K}^\times) \overset{\text{(OS)}}{\le} F_{>_\Omega 0}^\times \,\mathrm{nrd}(\widehat{K}^\times)\,\mathrm{nrd}(N_{\widehat{B}^\times}(\widehat{O})) \overset{m}{=} F_{>_\Omega 0}^\times \,\mathrm{nrd}(\widehat{E}) \overset{s}{\le} \widehat{F}^\times,$$

with total index at most 2.

By Proposition 31.4.4, we have that $\mathrm{Gen}\, O$ is optimally selective for S if and only if the right-most inequality (labelled s) is strict. Such an inequality is strict if and only if the total index is 2 and the left-most inequality is an equality. By Lemma 31.4.1, this happens if and only if the condition (OS) holds. This proves (a).

Statement (b) is a restatement of Corollary 31.4.6, and statement (c) follows from the second statement in Proposition 31.5.1. □

It has been a long and pretty technical road, so as refreshment we work through an example (cf. Maclachlan [Macl2008, §4, Example 1]).

Example 31.6.1. Let F be the totally real cubic field $\mathbb{Q}(b)$ where $b^3 - 4b - 1 = 0$; then F has discriminant 229 and ring of integers $R = \mathbb{Z}_F = \mathbb{Z}[b]$. The usual class group $\mathrm{Cl}\, R$ is trivial, but the narrow class group is $\mathrm{Cl}^+ R \simeq \mathbb{Z}/2\mathbb{Z}$, represented by the ideal $\mathfrak{p} = (b+1)\mathbb{Z}_F$ of norm 2—the ideal $\mathfrak{p}$ is principal, but there is no generator that is totally positive. The narrow class field $K = H^+ \supseteq F$ is quadratic, with $H^+ = F(\sqrt{b})$.

Let $B = \left(\dfrac{-1, b}{F}\right)$. Then b is positive at precisely one real place and negative at the other two, and $b \in \mathbb{Z}_F^\times$. Computing the Hilbert symbol at the even primes, we conclude that $\mathrm{Ram}(B)$ is equal to two real places. In particular, B is indefinite. The class group

$\mathrm{Cl}_{G(O)} R$ with modulus equal to these two real places is equal to $\mathrm{Cl}^+ R$, as we see by the real signs of b.

Next, we compute representatives of the type set of maximal orders for B. By strong approximation (Corollary 28.5.10), we have $\mathrm{Typ}\, O$ in bijection with $\mathrm{Cl}_{GN(O)} R$, so we need to compute the idelic normalizer: but B is unramified at all nonarchimedean places, and

$$N_{\widehat{B}^\times}(\widehat{O}) = N_{\mathrm{GL}_2(\widehat{F})}(\mathrm{M}_2(\widehat{R})) = \widehat{F}^\times \widehat{O}^\times.$$

Thus $\mathrm{nrd}(N_{\widehat{B}^\times}(\widehat{O})) = \widehat{F}^{\times 2}\widehat{R}^\times$, and

$$GN(O) = F^\times_{>_\Omega 0}\widehat{F}^{\times 2}\widehat{R}^\times = F^\times_{>_\Omega 0}\widehat{R}^\times = G(O).$$

In other words, the quotient map $\mathrm{Cl}_{G(O)} R \to \mathrm{Cl}_{GN(O)} R$ is an isomorphism, still a group of order 2. We conclude that $\#\,\mathrm{Typ}\, O = 2$.

We compute a maximal order

$$O = O_1 = \mathbb{Z}_F \oplus \mathbb{Z}_F i \oplus \mathbb{Z}_F \frac{b^2 i + j}{2} \oplus \mathbb{Z}_F \frac{b^2 + ij}{2}.$$

We conjugate this order by an ideal of reduced norm $\mathfrak{p}$ to get the second representative

$$O_2 = \mathbb{Z}_F \oplus \mathbb{Z}_F i \oplus \mathbb{Z}_F \frac{(b^2+b+1) + (b+1)i + j}{2} \oplus \mathbb{Z}_F \frac{(b+1) + (b^2+b+1)i + ij}{2}.$$

Therefore these orders represent the two types of maximal orders, and $\mathrm{Typ}\, O = \{[O_1], [O_2]\}$.

With all of these elements in place, we can observe selectivity (Main Theorem 31.1.7). We saw that both K and B are ramified at no nonarchimedean places and exactly the same set of real places. In particular, the field $K \hookrightarrow B$ embeds by the local-global principle. Let $S = \mathbb{Z}_K = \mathbb{Z}_F[w]$ be the maximal order in K. Then $w^2 - bw + 1 = 0$. Then $\mathrm{Emb}(\widehat{S}; \widehat{O}) \neq \emptyset$ (Proposition 30.5.3(a)).

The optimal selective condition (OS) holds because we took it so, $K = H^+$. It follows that S embeds in exactly one of O_1 or O_2. We find that

$$\alpha = \frac{b^2 + ij}{2b} \in O_1$$

satisfies $\alpha^2 - b\alpha + 1 = 0$ as desired; so S embeds in O_1 (and not O_2).

Remark 31.6.2. Without the hypothesis of strong approximation, it is very difficult to tease apart the contributions from different orders in the genus: indeed, the generating series for representation numbers for a definite quaternion order give coefficients of modular forms, discussed in Chapter 41.

31.7 $*$ Selectivity, without optimality

To conclude this chapter, we compare the above with a weaker condition than optimal selectivity, and close with connections to the literature. We continue notation and hypotheses from 31.1.4.

Definition 31.7.1. We say that $\operatorname{Gen} O$ is **selective** for S if there exists $O' \in \operatorname{Gen} O$ such that there is *no* embedding $S \hookrightarrow O'$ of R-algebras.

The difference between Definition 31.1.5 and Definition 31.7.1 is that in the latter, we do not insist that the embedding is optimal. It may happen that $\operatorname{Gen} O$ is selective for S, but $\operatorname{Gen} O$ is not *optimally* selective for S: such a situation arises exactly when there is an order $O' \in \operatorname{Gen} O$ such that S embeds in O' but does not optimally embed in O'.

Example 31.7.2. We return to Example 31.1.1. We saw that $\operatorname{Gen} O$ is optimally selective for the maximal order $\mathbb{Z}_K$. By Main Theorem 31.1.7, $\operatorname{Gen} O$ is also optimally selective for $S = \mathbb{Z}_F[\sqrt{-1}] \subseteq \mathbb{Z}_K$.

We claim that $\operatorname{Gen} O$ is *not* selective for S. For $O = \mathrm{M}_2(\mathbb{Z}_K)$, we take the normalized embedding

$$w \mapsto \alpha = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}.$$

The surprise is that we can also find an embedding $S \hookrightarrow O' = \begin{pmatrix} R & \mathfrak{p} \\ \mathfrak{p}^{-1} & R \end{pmatrix}$, just not an optimal one: we take

$$w \mapsto \alpha' = \begin{pmatrix} -\sqrt{-5} & 2 \\ 2 & \sqrt{-5} \end{pmatrix}. \tag{31.7.3}$$

(The argument given in Example 31.1.1 no longer applies, because the polynomial $x^2 - 1$ has nontrivial discriminant, giving just enough room for the prime $\mathfrak{p}$ to sneak in.)

For sanity (to show there is no contradiction with the main theorem of optimal selectivity), we confirm that the embedding (31.7.3) does not define an optimal embedding into O'. We have $2\mathbb{Z}_F = \mathfrak{p}^2$, so

$$\alpha' + 1 = \begin{pmatrix} 1 - \sqrt{-5} & 2 \\ 2 & 1 + \sqrt{-5} \end{pmatrix} \in \mathfrak{p}O'$$

so the order $R + \mathfrak{p}^{-1}(\alpha' + 1) \supseteq R + R\alpha' \simeq S$ embeds in O'.

31.7.4. If $\operatorname{Gen} O$ is optimally selective for S but not selective for S, then $S \subseteq O$ is optimal but there is an order O' such that $\phi' \colon S \hookrightarrow O'$ is an embedding but not an optimal embedding. Let $S' = \phi'(K) \cap O' \supsetneq S$. So there exists a prime $\mathfrak{p} \mid [S' : S]_R$, and in particular, S is not maximal at $\mathfrak{p}$. In particular, if S is integrally closed, then $\operatorname{Gen} O$ is selective for S if and only if $\operatorname{Gen} O$ is optimally selective for S.

Theorem 31.7.5 (Chinburg–Friedman, Chan–Xu, Guo–Qin). *Let O be an Eichler order of level $\mathfrak{M}$ and suppose that* $\operatorname{Gen} O$ *is optimally selective for S. Then* $\operatorname{Gen} O$ *is selective for S if and only if the following condition holds:*

(S) *If* $\mathfrak{p} \mid \operatorname{disc}_R S$ *and* $\operatorname{ord}_{\mathfrak{p}}(\mathfrak{M}) \neq \operatorname{ord}_{\mathfrak{p}}(\operatorname{disc}_R S)$, *then* $\mathfrak{p}$ *splits in* $K \supseteq F$.

Proof. Our very setup (section 31.3) is designed to count optimal embeddings, so to avoid lengthening this chapter, we refer the reader to Chinburg–Friedman [CF99, Theorem 3.3] for the case of maximal orders, and Chan–Xu [CX2004, Theorem 4.7] and Guo–Qin [GQ2004, Theorem 2.5] (independently) for Eichler orders. □

Remark 31.7.6. The condition Proposition 31.2.1(d) (one part of (OS) for Eichler orders) is not visible in the selectivity theorem for Eichler orders, but is implied by it: by (b), the extension $K \supseteq F$ is unramified so $\mathfrak{p}$ is unramified in K, thus $\operatorname{ord}_{\mathfrak{p}}(\operatorname{disc}_R(S))$ is even and necessarily not equal to $\operatorname{ord}_{\mathfrak{p}}(\mathfrak{M})$ if the latter is odd.

31.7.7. Chinburg–Friedman [CF99, Theorem 3.3] prove Theorem 31.7.5 for maximal orders, and they applied this theorem to embeddings in maximal arithmetic groups [CF99, Theorem 4.4]. Chinburg–Friedman proved their results in the language of the Bruhat–Tits tree of maximal orders. This selectivity theorem was then generalized to Eichler orders by Chan–Xu [CX2004, Theorem 4.7] and Guo–Qin [GQ2004, Theorem 2.5] (independently). Interestingly, while Guo–Qin follow Chinburg–Friedman in their proof, Chan–Xu instead use results on exceptional spinor genera and their results are phrased and proven in the language of indefinite integral quadratic forms. (These results are given for number fields, but the proofs adapt to global fields as pursued here.)

Some selectivity theorems beyond those for Eichler orders are also known. Arenas-Carmona [A-C2013, Theorem 1.2] considers more general intersections of maximal orders. Linowitz [Lin2012, Theorems 1.3–1.4] gives a selectivity theorem for (optimal) embeddings into arbitrary orders, subject to some additional technical (coprimality) hypotheses. More generally, selectivity theorem have been pursued in the more context of central simple algebras: see e.g. Linowitz–Shemanske [LS2012] and Arenas-Carmona [A-C2012].

However, these selectivity results do not prove Main Theorem 31.1.7 on the nose, either because they deal with selectivity instead of optimal selectivity or do not prove the more powerful statement that the embedding numbers are in fact equal. On the latter point, a general setup to establish equality of embedding numbers can be found in work of Linowitz–Voight [LV2015, §2].

31.8 ∗ Isospectral, nonisometric manifolds

We conclude with an application to geometry. We need to borrow from the future, so the reader is invited to read this section lightly, using it as present and future motivation; and then to return to this section after a more careful reading of Chapter 38.

In 1966, Kac [Kac66] famously asked: "Can one hear the shape of a drum?" Put another way, if you know the frequencies at which a drum vibrates, can you determine its shape? This beautiful question has led to an almost countless number of articles: see Giraud–Thas [GT2010] for a survey.

To restate the question in a mathematical framework, let M be a connected, compact Riemannian manifold. Associated to M is the **Laplace operator**, defined by $\Delta(f) := -\operatorname{div}(\operatorname{grad}(f))$ for $f \in L^2(M)$ a square-integrable function on M. The eigenvalues of Δ on the space $L^2(M)$ form an infinite, discrete sequence of nonnegative real numbers

$0 = \lambda_0 < \lambda_1 \leq \lambda_2 \leq \dots$, called the **spectrum** of M. By the study of the wave equation, the spectrum of M can essentially be thought of as the frequencies produced by a drum "in the shape of M". Two Riemannian manifolds are said to be **Laplace isospectral** if they have the same spectra. We are led to ask: if two Riemannian manifolds are Laplace isospectral, are they isometric?

A general algebraic method for constructing Laplace isospectral manifolds is due to Sunada [Sun85] (see the surveys by Gordon [Gor2000, Gor2009]), arising from *almost conjugate* subgroups of a finite group of isometries acting on a manifold—providing a negative answer to Kac's question by Gordon–Webb–Wolpert [GWW92]. But preceding Sunada, in 1980 Vignéras [Vig80b] constructed such examples—indeed, one of the applications in her book on quaternion algebras [Vig80a] was to explain this construction!

Theorem 31.8.1 (Vignéras). *For every $m \geq 2$, there exist Laplace isospectral and nonisometric manifolds of dimension m.*

We sketch a proof of this theorem in this section, with attention to the particular detail of selectivity. In 1994, Maclachlan–Rosenberger [MacRos94] claimed to have produced a pair of Laplace isospectral, nonisometric hyperbolic 2-orbifolds of genus 0, but then Buser–Flach–Semmler [BFS2008] later showed that these examples were too good to be true! The subtle issue they found: the phenomenon of selectivity.

Our construction is quaternionic, of course. We consider the situation of sections 38.2–38.3, specifically the setup in 38.2.1 and 38.3.1. Let F be a number field with r real places and c complex places, so that $[F : \mathbb{Q}] = r + 2c = n$. Let B be a quaternion algebra over F and suppose that B is split at t real places. We have an embedding (38.2.2) $\iota\colon B \hookrightarrow \mathrm{M}_2(\mathbb{R})^t \times \mathrm{M}_2(\mathbb{C})^c$. Letting $\mathcal{H} := (\mathbf{H}^2)^t \times (\mathbf{H}^3)^c$ as in (38.2.9) and

$$\mathrm{P}B^\times_{>0} := B^\times_{>0}/F^\times,$$

we see that $\iota(\mathrm{P}B^\times_{>0}) \leq \mathrm{Isom}^+(\mathcal{H})$ acts on $\mathcal{H}$ by linear fractional transformations as orientation-preserving isometries.

Let $R = \mathbb{Z}_F$ be the ring of integers of F and let $O \subset B$ be an R-order. Then the group $\Gamma^1(O) := \iota(O^1/\{\pm 1\}) \leq \mathrm{P}B^\times_{>0}$ is a discrete subgroup acting properly on $\mathcal{H}$ by isometries. Suppose now that B is a division algebra, and further that the only elements of O^1 of finite order are ± 1. (We will soon see that this assumption can be satisfied; in some sense, it is quite typical.) Let $X^1(O) := \Gamma^1(O)\backslash\mathcal{H}$ be the quotient, a Riemannian manifold of dimension $m := 2t + 3c$.

The condition of Laplace isospectrality is implied by the following arithmetic condition in terms of conjugacy classes of embeddings.

Theorem 31.8.2. *Let $O' \in \mathrm{Gen}\, O$. Suppose that for every quadratic field $K \supseteq F$ and every quadratic R-order $S \subseteq K$, we have the equality of embedding numbers*

$$m(S, O; O^1) = m(S, O'; O'^1). \tag{31.8.3}$$

Then $X^1(O)$ and $X^1(O')$ are Laplace isospectral.

Proof. See Vignéras [Vig80b, Théorème 6]. The statement there is in terms of all embeddings, not just optimal embeddings—but the total count of conjugacy classes of embeddings of an quadratic order is a sum of the corresponding count of optimal embeddings of superorders (as in 31.7), so it is sufficient to have a genial order.

The key ingredient in the proof is the *Selberg trace formula*, which allows us to show that the spectra of the Laplace operators agree by the stronger condition of representation equivalence: for a more general point of view on this deduction, see Deturck–Gordon [DG89]. □

The rub is in the equality (31.8.3). The restriction of the equivalence classes to units of reduced norm 1 is harmless: by Lemma 30.3.14, we have $m(S, O; O^1) = m(S, O; O^\times)[\mathrm{nrd}(O^\times) : \mathrm{nrd}(S^\times)]$, and as a consequence of strong approximation we have $[\mathrm{nrd}(O^\times) : \mathrm{nrd}(S^\times)] = [\mathrm{nrd}(\widehat{O}^\times : \mathrm{nrd}(\widehat{S}^\times)]$. Thus if $O' \in \mathrm{Typ}\, O$, then (31.8.3) holds if and only if $m(S, O; O^\times) = m(S, O'; O'^\times)$.

Finally, selectivity enters! We suppose that O is an Eichler order. By Main Theorem 31.1.7(c), we have the desired equality when $\mathrm{Gen}\, O$ is genial (i.e., not optimally selective)—and by part (b), this equality may fail for some S.

Corollary 31.8.4. *Suppose that* $\mathrm{Gen}\, O$ *is genial and* $O' \in \mathrm{Gen}\, O$ *has* $O' \not\simeq O$. *Suppose further that* $\sigma(\mathrm{Ram}(B)) \neq \mathrm{Ram}(B)$ *for all* $\sigma \in \mathrm{Aut}(F)$. *Then* $X^1(O)$ *and* $X^1(O')$ *are Laplace isospectral, nonisometric Riemannian manifolds of dimension m.*

Proof. Isospectrality follows from Theorem 31.8.2 with Main Theorem 31.1.7. To show that X, X' are not isometric, since $O' \not\simeq O$ we know that $\mathrm{P}O'$ is not conjugate to $\mathrm{P}O$ in $B^\times$, but we need this for the groups $\Gamma^1(O'), \Gamma^1(O)$ in $\mathrm{Isom}^+(\mathcal{H})$, which is slightly larger (see Remark 38.2.11). We leave the details to Exercise 31.3 (or see Linowitz–Voight [LV2015, Proposition 2.24]). □

The remainder of the proof of Theorem 31.8.1 involves finding suitable data F, B, O. We exhibit a pair for $m = 2$ following Linowitz–Voight [LV2015, Example 5.2], giving a pair of compact hyperbolic 2-manifolds of genus 6 which are isospectral but not isometric.

Example 31.8.5. Let $F = \mathbb{Q}(w)$ where w is a root of the polynomial $x^4-5x^2-2x+1=0$. Then F is a totally real quartic field with discriminant $d_F = 5744 = 2^4 359$, Galois group S_4, class number $\#\mathrm{Cl}\, R = 1$ and narrow class number $\#\mathrm{Cl}^+ R = 2$. Let B be the quaternion algebra over F which is ramified at the prime ideal $\mathfrak{p}_{13}$ of norm 13 generated by $b := w^3 - w^2 - 4w$ and three of the four real places, with split place $w \mapsto -0.751024\ldots$: then $B = \left(\dfrac{a, b}{F}\right)$ where $a = w^3 - w^2 - 3w - 1$ is a root of $x^4 + 8x^3 + 12x^2 - 1$.

A maximal order $O \subset B$ is given by

$$O = R \oplus Ri \oplus \frac{(w^3+1) + w^2 i + j}{2} R \oplus \frac{(w+1) + (w^3+1)i + ij}{2} R;$$

O has type number 2, so there exists two isomorphism classes of maximal orders $O_1 = O$ and O_2.

We claim that O_1 and O_2 have no elements of finite order other than ± 1. Indeed, if we had such an element of order q then $F(\zeta_{2q})$ is a cyclotomic quadratic extension of F, whence $\mathbb{Q}(\zeta_{2q})^+ \subseteq F$; but F is primitive, so the only cyclotomic quadratic extensions of F are $K = F(\sqrt{-1})$ and $K = F(\sqrt{-3})$. But as $\mathfrak{p}_{13}$ splits completely in $F(\sqrt{-1})$ and $F(\sqrt{-3})$, neither field embeds into B. We conclude that the groups Γ_i are torsion free.

Since B is ramified at a finite place, the genus of O is genial by Proposition 31.2.1, and since $\operatorname{Aut}(F)$ is trivial, the hypothesis of Corollary 31.8.4 are satisfied: X_1^1, X_2^1 are Laplace isospectral, but not isometric.

Finally, by Theorem 39.1.13, we have $\operatorname{area}(X_i) = 20\pi$, so $g(X_i) = 6$ for $i = 1, 2$. Fundamental domains for these are given in Figure 31.8.6.

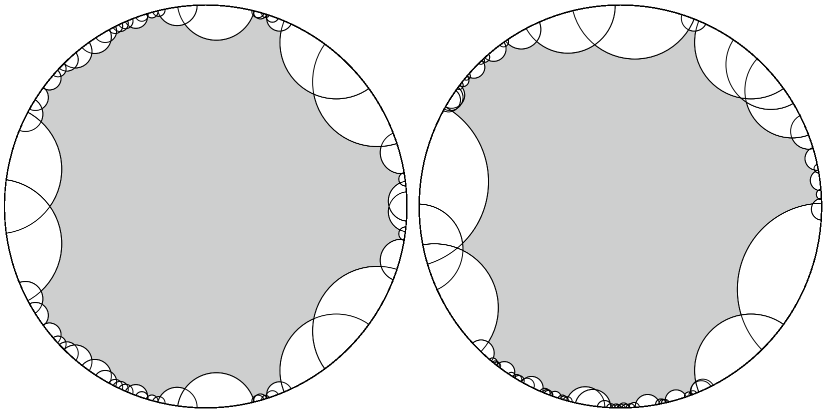


Figure 31.8.6: Fundamental domain for the genus 6 manifolds $X(\Gamma_1^1)$,$X(\Gamma_2^1)$

We obtain a second example by choosing the split real place $w \mapsto -1.9202\dots$, and since F is not Galois, as in the case of the 2-orbifold pairs 2 and 3, these are pairwise nonisometric.

For an example with $m = 3$, see Exercise 31.4.

Exercises

▹ 1. Prove Main Theorem 31.1.7 in the case $K \simeq F \times F$: to be precise, show that an R-order $S \subseteq F \times F$ embeds equally in all Eichler R-orders. *[Hint: we must have $B \simeq \mathrm{M}_2(F)$, so reduce to the case where S is embedded in the diagonal and then conjugate.]*

2. The following exercise gives insight into the proof of Theorem 31.7.5 on selectivity. Let R be local, and let O be an Eichler order of level $\mathfrak{p}^e$.

   Let $\phi\colon S \hookrightarrow O$ be an optimal embedding that is normalized and associated to $x \in R$, so represented by

$$\alpha = \begin{pmatrix} x & 1 \\ -f_\gamma(x) & t - x \end{pmatrix}$$

as in Definition 30.6.8.

(a) Compute $\nu^{-1}\alpha\nu$ for the matrix
$$\nu = \begin{pmatrix} a & b \\ c & d \end{pmatrix}.$$

(b) Show that the off-diagonal entries are equal to
$$(ad-bc)^{-1}\operatorname{Nm}_{K|F}(b\gamma - (bx+d)),$$
$$(ad-bc)^{-1}\operatorname{Nm}_{K|F}(a\gamma - (ax+c))$$
so belong to $(\det\nu)^{-1}\operatorname{Nm}_{K|F}(K^\times)$.

(c) Suppose $\alpha' = \nu^{-1}\alpha\nu$ gives a normalized, optimal embedding. Show that if $\mathfrak{p}$ is inert in $K \supseteq F$, then $\det\nu$ has even valuation.

▷ 3. With notation as in section 31.8, let $\Gamma, \Gamma' \leq B^\times/F^\times$ and $\iota\colon B^\times_{>0}/F^\times \hookrightarrow \operatorname{Isom}^+(\mathcal{H})$. Then $X(\Gamma), X(\Gamma')$ are isometric if and only if there exists a permutation σ of the factors of $G := \operatorname{Isom}^+(\mathcal{H})$ and $\nu \in \iota(B^\times)$ such that $\sigma(\Gamma) = \nu\Gamma'\nu^{-1}$ and a $\mathbb{Q}$-algebra automorphism $\tau\colon B \to B$ such that the diagram
$$\begin{array}{ccccc} B^\times & & \xrightarrow{\quad\tau\quad} & & B^\times \\ \downarrow\iota & & & & \downarrow\iota \\ G & \xrightarrow{\sigma} & G & \xrightarrow{\nu} & G \end{array}$$
commutes, where ν acts on G by conjugation. [Hint: use the Skolem–Noether theorem.]

4. In this exercise, we exhibit Laplace isospectral, nonisometric hyperbolic 3-manifolds, following Linowitz–Voight [LV2015, Example 6.3]. Let $F = \mathbb{Q}(w)$ be the quintic field with discriminant -43535 and defining polynomial $x^5 - x^4 + 3x^3 - 3x + 1$. Let $B := \left(\dfrac{3w^3-2, -13}{F}\right)$.

(a) Show that B is ramified at the three real places of F and the prime ideal $\mathfrak{p} = (w^4 - w^3 + 3w^2 - w - 2)$ of norm 13.
(b) Let O be a maximal order in B. Show that the type number of O is equal to 2.
(c) Let $O = O_1$ and O_2 be representatives of $\operatorname{Typ} O$. Show that the only elements of finite order in O_i^1 are ± 1.
(d) Show that $\operatorname{Gen} O$ is genial.
(e) Let $X_i^1 := \Gamma^1(O_i)\backslash\mathbf{H}^3$, with notation as in section 31.8. Show that X_1^1 and X_2^1 are Laplace isospectral, nonisometric 3-manifolds.
(f) Show that $\operatorname{vol}(X_1^1) = \operatorname{vol}(X_2^1) = 51.024566\ldots$

This chapter is licensed under the terms of the Creative Commons Attribution-NonCommercial 4.0 International License (http://creativecommons.org/licenses/by-nc/4.0/), which permits any noncommercial use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license and indicate if changes were made.

The images or other third party material in this chapter are included in the chapter's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the chapter's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.

Part IV
Geometry and topology

Chapter 32

Unit groups

Having moved from algebra and arithmetic to analysis, and in particular the study of class numbers, in this part we consider geometric aspects of quaternion algebras, and the unit group of a quaternion order acting by isometries on a homogeneous space.

32.1 ▹ Quaternion unit groups

By way of analogy, we consider what happens for quadratic orders. In this case, just as with class groups, the behavior of unit groups is quite different depending on whether the asociated quadratic field K is real or imaginary.

In the imaginary case, the unit group is finite, as the norm equation $\operatorname{Nm}_{K|\mathbb{Q}}(\gamma) = 1$ has only finitely many solutions for integral γ: these are elements of a 2-dimensional lattice in $\mathbb{C}$ that lie on the unit circle. Such an element is a root of unity that satisfies a quadratic equation over $\mathbb{Q}$, and so only two imaginary quadratic orders having units other than ± 1 are the Gaussian order $\mathbb{Z}[i]$ of discriminant -4 and the Eisenstein order $\mathbb{Z}[\rho]$ with $\rho := (-1 + \sqrt{-3})/2$ of discriminant -3: see Figure 32.1.1.

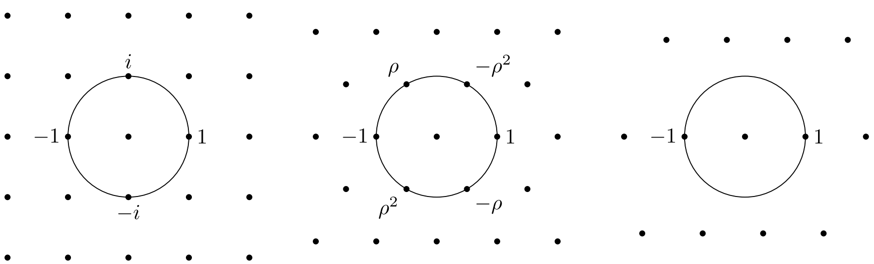


Figure 32.1.1: Units in imaginary quadratic orders

Orders O in a *definite* quaternion algebra B over $\mathbb{Q}$ behave like orders in an imaginary quadratic field. The unit group of such an order is finite, as the solutions to $\operatorname{nrd}(\alpha) = 1$ with $\alpha \in O$ are elements of a 4-dimensional lattice in $\mathbb{R}^4$ again with bounded size. In section 11.5, after a close investigation of the case of Hurwitz units,

© The Author(s) 2021

J. Voight, *Quaternion Algebras*, Graduate Texts in Mathematics 288,
https://doi.org/10.1007/978-3-030-56694-4_32

we classified the possibilities, embedding $\mathcal{O}^\times/\{\pm 1\} \hookrightarrow \mathbb{H}^1/\{\pm 1\} \simeq \mathrm{SO}(3)$ as a finite rotation group: the groups $\mathcal{O}^\times$ that arise over $\mathbb{Q}$ are either cyclic of order $2, 4, 6$, quaternion Q_8 of order 8, binary dihedral $2D_6$ of order 12, or the binary tetrahedral group $2T$ of order 24. In this chapter, we take up this task in the context of a general definite quaternion order, and realize all finite rotation groups using quaternions.

Now we turn to real quadratic fields and correspondingly indefinite quaternion algebras. For the real quadratic order $\mathbb{Z}[\sqrt{d}]$ with $d > 0$, the units are solutions to the Pell equation $\mathrm{Nm}_{K|\mathbb{Q}}(x - y\sqrt{d}) = x^2 - dy^2 = \pm 1$ with $x, y \in \mathbb{Z}$. All solutions up to sign are given by powers of a **fundamental solution** which can be computed explicitly using continued fractions; consequently, $\mathbb{Z}[\sqrt{d}]^\times = \langle -1, u\rangle \simeq \mathbb{Z}/2\mathbb{Z} \times \mathbb{Z}$ where $u = x + y\sqrt{d}$ is the **fundamental unit**. The fundamental unit often (but not always) has large height (in the sense that $x, y \in \mathbb{Z}$ are large in absolute value), being of exponential size in the discriminant, by theorems of Schur and Siegel. The unit group of the ring of integers of $\mathbb{Q}(\sqrt{d})$ for $d \equiv 1 \pmod 4$ is treated in a similar way, by considering the norm equation $x^2 - xy + cy^2 = \pm 1$ where $c = (1 - d)/4$.

For quaternions, we are led to consider units in the standard order

$$\mathcal{O} := \mathbb{Z} \oplus \mathbb{Z}i \oplus \mathbb{Z}j \oplus \mathbb{Z}ij \subseteq B := \left(\frac{a, b}{\mathbb{Q}}\right)$$

in an indefinite quaternion algebra, with $a, b \in \mathbb{Z}$ and say $a > 0$. The norm condition $\mathrm{nrd}(\gamma) = \pm 1$ for $\gamma = t + xi + yj + zij$ then reads

$$t^2 - ax^2 - by^2 + abz^2 = \pm 1 \tag{32.1.2}$$

with $t, x, y, z \in \mathbb{Z}$. Amusingly, this "quaternion Pell equation" includes the Pell equation for $\mathbb{Z}[\sqrt{a}]$ by setting $y = z = 0$, and in fact by considering embeddings of quadratic orders (the subject of Chapter 30), we see that this equation combines *all* Pell equations satisfying certain congruence conditions. Combining these Pell equations, we see that the group of solutions is an infinite, noncommutative group. (The case of an order different from the standard one will give a different norm equation, but the same conclusions.) See Jahangiri [Jah2010] for a Diophantine interpretation of the structure of the unit group of a quaternion order as a quaternionic Pell equation.

We will seek to understand the group $\mathcal{O}^\times$ by its action on a suitable space, and in this way we are led to consider groups acting discretely on symmetric spaces; we will discover that the group $\mathcal{O}^\times$ is finitely presented and in particular finitely generated, so we still can think of a set of *fundamental solutions* (given by generators) whose products generate all solutions to (32.1.2). For example, we may take $\mathcal{O} = \mathrm{M}_2(\mathbb{Z}) \subseteq \mathrm{M}_2(\mathbb{Q})$, where $\mathcal{O}^\times = \mathrm{GL}_2(\mathbb{Z})$, generated by the elementary matrices $\begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}, \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$. Accordingly, our investigation is detailed but fruitful, involving the theory of Fuchsian and Kleinian groups.

In this chapter, we begin by discussing the general structure of these groups.

32.2 Structure of units

Throughout this chapter, we use the following notation, recalling our notation for global fields from section 14.4. Let F be a global field, let $S \subseteq \mathrm{Pl}(F)$ be an eligible set of

places of F, and let $R = R_S$ be the global ring associated to S, the ring of S-integers of F. (As always, the reader may keep the case $F = \mathbb{Q}$, $S = \{\infty\}$, and $R = \mathbb{Z}$ in mind.) Further, let B be a quaternion algebra over F, and let $O \subset B$ be an R-order.

We are interested in the structure of the group $O^\times$. Since $Z(B^\times) = F^\times$, we have $R^\times \leq Z(O^\times)$ central. We understand the structure of $R^\times$ by Dirichlet's unit theorem, as follows.

32.2.1. From Dirichlet's unit theorem (and its extension to S-units and the function field case), the group $R^\times$ of units is a finitely generated abelian group of rank $\#S - 1$, so that

$$R^\times \simeq \mathbb{Z}/w\mathbb{Z} \oplus \mathbb{Z}^{\#S-1} \tag{32.2.2}$$

where w is the number of roots of unity in F. (The proof is briefly recalled in 32.3.1.)

The group $O^\times$ is (in general) noncommutative, so we should not expect a description like 32.2.1. But to get started, we consider the quotient $O^\times/R^\times$, and the reduced norm map which gets us back into $R^\times$.

32.2.3. We recall the theorem on norms (see section 14.7): as before, let

$$\Omega := \{v \in \operatorname{Ram} B : v \text{ real}\} \subseteq \operatorname{Pl} F \tag{32.2.4}$$

be the set of real ramified places in B (recalling that complex places cannot be ramified), and

$$F^\times_{>_\Omega 0} := \{x \in F^\times : v(x) > 0 \text{ for all } v \in \Omega\} \tag{32.2.5}$$

the set of elements that are positive at the places $v \in \Omega$. If F is a function field, then $\Omega = \emptyset$, and $R^\times_{>_\Omega 0} = R^\times$. The Hasse–Schilling norm theorem (Main Theorem 14.7.4) says that $\operatorname{nrd}(B^\times) = F^\times_{>_\Omega 0}$. Letting

$$R^\times_{>_\Omega 0} := R^\times \cap F^\times_{>_\Omega 0}, \tag{32.2.6}$$

we conclude that $\operatorname{nrd}(O^\times) \leq R^\times_{>_\Omega 0}$. (Strictly speaking, we only needed the containment $\operatorname{nrd}(B^\times) \leq F^\times_{>_\Omega 0}$ which follows directly from local considerations; but we pursue finer questions below.)

32.2.7. In light of 32.2.3, the reduced norm gives an exact sequence

$$1 \to O^1 \to O^\times \xrightarrow{\operatorname{nrd}} R^\times_{>_\Omega 0} \tag{32.2.8}$$

where

$$O^1 := \{\gamma \in O^\times : \operatorname{nrd}(\gamma) = 1\} \tag{32.2.9}$$

is the subgroup of units of reduced norm 1.

Since $\operatorname{nrd}(R^\times) = R^{\times 2}$ by the squaring map, we have $O^1 \cap R^\times = \{\pm 1\}$, so (32.2.8) yields

$$1 \to \frac{O^1}{\{\pm 1\}} \to \frac{O^\times}{R^\times} \xrightarrow{\operatorname{nrd}} \frac{R^\times_{>_\Omega 0}}{R^{\times 2}} \tag{32.2.10}$$

Since the group $R^\times$ is finitely generated, the group $R^\times_{>_\Omega 0}/R^{\times 2}$ is a finite, elementary abelian 2-group. (In general, the reduced norm (the last) map in (32.2.10) need not be surjective.)

32.2.11. In general, the exact sequence (32.2.10) does not split, so the group $O^\times/R^\times$ will be a nontrivial extension of $O^1/\{\pm 1\}$ by an elementary abelian 2-group.

Example 32.2.12. If $B = \mathrm{M}_2(F)$ and $O = \mathrm{M}_2(R)$, then $O^\times = \mathrm{GL}_2(R)$ and $O^\times/R^\times = \mathrm{GL}_2(R)/R^\times =: \mathrm{PGL}_2(R)$. The reduced norm is the determinant, which is surjective, and so the exact sequence (32.2.10) can be extended to

$$1 \to \mathrm{PSL}_2(R) \to \mathrm{PGL}_2(R) \xrightarrow{\det} R^\times/R^{\times 2} \to 1.$$

32.2.13. In light of Example 32.2.12, it is natural to write $\mathrm{P}O^\times := O^\times/R^\times$ and $\mathrm{P}O^1 := O^1/\{\pm 1\}$.

Remark 32.2.14. Some authors write $\mathrm{GL}_1(O) = O^\times$ and $\mathrm{SL}_1(O) = O^1$, and this notation suggests generalizations. In such situations, it is natural to write $\mathrm{PGL}_1(O) = O^\times/R^\times$ and $\mathrm{PSL}_1(O) = O^1/\{\pm 1\}$.

32.2.15. Suppose F is a number field. Then the group $R^\times_{>_\Omega 0}/R^{\times 2}$ is canonically isomorphic to a quotient of class groups, as follows. Let $\mathbb{Z}_F$ be the ring of integers of F, and let $\mathrm{Cl}_\Omega \mathbb{Z}_F$ denotes the class group of F given by the quotient of the group of fractional ideals by the subgroup of principal ideals with a generator positive at all places in Ω. Then $R^\times_{>_\Omega 0}/R^{\times 2}$ is isomorphic to the quotient of $\mathrm{Cl}_\Omega \mathbb{Z}_F$ by $\mathrm{Cl}\,\mathbb{Z}_F$ and the group generated by the finite primes in S.

32.3 Units in definite quaternion orders

In this section, we show that for a definite quaternion order O, the scalar units $R^\times$ are of finite index in $O^\times$—i.e., the group $\mathrm{P}O^\times = O^\times/R^\times$ is a finite group.

32.3.1. To build intuition, suppose F is a number field with r real places and c complex places. Recall the proof of Dirichlet's unit theorem: we define a map

$$\begin{aligned} R^\times &\to \mathbb{R}_S \\ x &\mapsto \left(m_v \log|x|_v\right)_v . \end{aligned} \tag{32.3.2}$$

The kernel of this map is the group of roots of unity (the torsion subgroup of $R^\times$). The image lies inside the trace zero hyperplane $\sum_{v\in S} x_v = 0$ by the product formula (14.4.6), and it is discrete and cocompact inside this hyperplane, so it is isomorphic to $\mathbb{Z}^{S-1}$. In particular, $R^\times$ is finite if and only if $\#S = 1$; since S always contains the set of archimedean places of size $r + c$, we see that $R^\times$ is finite if and only if $(r, c) = (1, 0), (0, 1)$, so $F = \mathbb{Q}$ or F is an imaginary quadratic field.

Remark 32.3.3. Informally, one might say that $R^\times$ is finite only when the completions at the places in S provide "no room" for the unit group to become infinite. This is analogous to the informal case for strong approximation in 28.5.4: if there is a place $v \in S$ where B^1_v is not compact, then there is enough room for B^1 to "spread out" and become dense.

32.3.4. Recall (Definition 28.5.1) that B is S-definite if $S \subseteq \mathrm{Ram}(B)$, i.e., every place in S is ramified in B. In particular, if F is a number field, then since a complex place is split and S contains the archimedean places, if B is S-definite then F is totally real; in this case, when S is exactly the set of archimedean places, we simply say that B is definite.

32.3.5. Consider the setup in analogy with Dirichlet's unit theorem 32.3.1. We consider the embedding of B into the completions at all places in S:

$$B \hookrightarrow B_S := \prod_{v \in S} B_v.$$

By Exercise 27.14, R is discrete in $F_S = \prod_{v \in S} F_v$ and O is discrete in B_S (the point being that in the number field case, S contains all archimedean places). Consequently, the injections

$$\begin{aligned} O^\times/R^\times &\hookrightarrow (B_S)^\times/(F_S)^\times := \prod_{v \in S} B_v^\times/F_v^\times \\ O^1 &\hookrightarrow (B_S)^1 = \prod_{v \in S} B_v^1 \end{aligned} \tag{32.3.6}$$

have discrete image.

Depending on whether the place v is nonarchimedean (split or ramified) or archimedean (split real, ramified real, or complex), we have a different target component $B_v^\times/F_v^\times$ or B_v^1. The major task of Part IV is to describe these possibilities in detail and look at the associated symmetric spaces.

We begin with the simplest case, where the unit groups involved are finite.

Proposition 32.3.7. *The group $O^\times/R^\times$ is finite if and only if O^1 is finite if and only if B is S-definite.*

Proof. By the exact sequence (32.2.10), the group $O^\times/R^\times$ is finite if and only if the group O^1 is finite.

First, suppose that B is S-definite. Then by definition, for each $v \in S$, the completion B_v is a division algebra over F_v. But each B_v^1 is compact, from the topological discussion in section 13.5. Therefore in (32.3.6), the group O^1 is a closed, discrete subgroup of a compact group—hence finite.

Now suppose B is not S-definite. Then there is a place $v_0 \in S$ that is unramified; we will correspondingly find an element of infinite order (like solutions to the quaternion Pell equation coming from the original Pell's equation (32.1.2)). We have $B_{v_0} \simeq \mathrm{M}_2(F_{v_0})$, so there exists $\alpha \in B$ be such that the reduced characteristic polynomial splits in F_{v_0}; we may suppose without loss of generality that $K = F[\alpha]$ is a field. Let S be the integral closure of R in K (not to be confused with the set S). Then by the Dirichlet S-unit theorem (32.2.1), the rank of $S^\times/R^\times$ is at least 1: the set of places $w \in \mathrm{Pl}(K)$ such that w lies above $v \in S$ contains at least one element from each v and two above v_0, because it is split. So there is an element $\gamma \in S^\times/R^\times$ of infinite order. As R-lattices, the order $S \cap O$ has finite R-index and hence finite index in S, so $S^\times/(S \cap O)^\times$ is a finite group, and therefore a sufficiently high power of γ lies in $(S \cap O)^\times \subseteq O^\times$, and $O^\times$ is infinite. ∎

Example 32.3.8. Let $B = (-1, -1 \mid \mathbb{Q})$ and let O be the $\mathbb{Z}$-order generated by i, j, so that $S = \{\infty\}$. Then B is S-definite, and $O^\times = \langle i, j\rangle \simeq Q_8$ is the quaternion group of order 8.

Now consider $S = \{2, \infty\}$; then B is still S-definite. We find

$$\begin{aligned} O[1/2]^\times &= \langle 2, i, j, 1+i\rangle \\ O[1/2]^\times/\mathbb{Z}[1/2]^\times &= \langle 2, i, j, 1+i\rangle/\langle -1, 2\rangle \simeq Q_8 \rtimes \mathbb{Z}/2\mathbb{Z} \end{aligned} \tag{32.3.9}$$

(Exercise 32.2).

Finally, if we take $S = \{5, \infty\}$, then B is no longer S-definite; and $O[1/5]^\times$ contains the element $2+i$ of norm $5 \in \mathbb{Z}[1/5]^\times$ and infinite order.

32.3.10. Suppose B is S-definite. Then by Proposition 32.3.7, the group $O^\times/R^\times$ is finite. Since we have an embedding

$$O^\times/R^\times \hookrightarrow B^\times/F^\times$$

it follows that a $O^\times/R^\times$ is a finite subgroup of $\mathrm{P}B^\times$, so a classification of finite subgroups of $\mathrm{P}B^\times$ gives a list of possible definite unit groups; we make this our task in the remainder of this chapter.

32.4 Finite subgroups of quaternion unit groups

We now embark on a classification of finite subgroups of $\mathrm{P}B^\times = B^\times/F^\times$ and $\mathrm{P}B^1 = B^1/\{\pm 1\}$; this is akin to first getting acquainted with the roots of unity in a number field. Suppose throughout the rest of this chapter that F is a number field; we allow B to be definite or indefinite.

We begin in this section with the classification of the possible groups up to isomorphism that goes back at least to Klein [Kle56, Chapter II]: the original book dates back to 1884 and is undoubtedly one of the most influential books of 19th century mathematics. See also the descriptions by Coxeter [Coxtr40] and Lamotke [Lamo86, Chapters I–II] for a presentation of the regular solids, finite rotation groups, as well as finite subgroups of $\mathrm{SL}_2(\mathbb{C})$.

Proposition 32.4.1. *Let $\Gamma < \mathrm{P}B^\times$ be a finite group. Then Γ is cyclic, dihedral, or an exceptional group A_4, S_4, A_5.*

We met these groups already in Proposition 11.5.2, and the proof is an extension of this result.

Proof. Let v be an archimedean place of F. Then the natural map $B^\times \to B_v^\times/F_v^\times$ has kernel $F_v^\times \cap B^\times = F^\times$, so the group homomorphism $B^\times/F^\times \hookrightarrow B_v^\times/F_v^\times$ is injective.

First suppose that v is a ramified (real) place, so $B_v \simeq \mathbb{H}$ and

$$B_v^\times/F_v^\times \simeq \mathbb{H}^\times/\mathbb{R}^\times \simeq \mathbb{H}^1/\{\pm 1\}.$$

By Corollary 2.4.21, we have $\mathbb{H}^1/\{\pm 1\} \simeq \mathrm{SO}(3)$ so Γ is a finite rotation group: these are classified in Proposition 11.5.2.

In general, we seek to conjugate the group Γ in order to reduce to the case above. We may prove the lemma after making a base extension of F, so we may suppose that v is complex, with $B_v \simeq \mathrm{M}_2(\mathbb{C})$. Then $B_v^\times/F_v^\times \simeq \mathrm{PGL}_2(\mathbb{C})$, and via the injection $B^\times/F^\times \hookrightarrow \mathrm{PGL}_2(\mathbb{C})$ we obtain a finite subgroup $\Gamma \subseteq \mathrm{PGL}_2(\mathbb{C})$. The natural map $\mathrm{SL}_2(\mathbb{C}) \to \mathrm{PGL}_2(\mathbb{C})$ is surjective, as we may rescale every invertible matrix by a square root of its determinant to have determinant 1, and its kernel is $\{\pm 1\}$, giving an isomorphism $\mathrm{PSL}_2(\mathbb{C}) \simeq \mathrm{PGL}_2(\mathbb{C})$. We then lift Γ under the projection $\mathrm{SL}_2(\mathbb{C}) \to \mathrm{PSL}_2(\mathbb{C})$ to a finite group (containing -1). We have

$$\mathbb{H}^1 \simeq \mathrm{SU}(2) = \{A \in \mathrm{SL}_2(\mathbb{C}) : A^*A = 1\} \hookrightarrow \mathrm{SL}_2(\mathbb{C}) \tag{32.4.2}$$

as in 2.4.2. If $\langle\ ,\ \rangle$ denotes the canonical (Hermitian) inner product on $\mathbb{C}^2$ defined by $\langle z, w\rangle = z^*w$ (as column vectors), then SU(2) is precisely the group of matrices of determinant 1 preserving $\langle\ ,\ \rangle$, i.e.,

$$\mathrm{SU}(2) = \{A \in \mathrm{SL}_2(\mathbb{C}) : \langle Az, Aw\rangle = \langle z, w\rangle \text{ for all } z, w \in \mathbb{C}^2\}$$

since $\langle Az, Aw\rangle = z^*(A^*A)w = z^*w$ if and only if $A^*A = 1$. We now define a Γ-invariant Hermitian inner product on $\mathbb{C}^2$ by averaging: for $z, w \in \mathbb{C}^2$, we define

$$\langle z, w\rangle_\Gamma := \frac{1}{\#\Gamma}\sum_{\gamma\in\Gamma}\langle \gamma z, \gamma w\rangle.$$

Choose an orthonormal basis for $\langle\ ,\ \rangle_\Gamma$ and let $T \in \mathrm{SL}_2(\mathbb{C})$ be the change of basis matrix relative to the standard basis. Then $\langle z, w\rangle_\Gamma = \langle Tz, Tw\rangle$ and therefore $T\Gamma T^{-1} \subset \mathrm{SU}(2)$. The result now follows from (32.4.2) and the previous case. □

32.5 Cyclic subgroups

In the next few sections, we discuss each of the possibilities in Proposition 32.4.1 in turn, following Chinburg–Friedman [CF2000]. We begin with cyclic subgroups.

There are always many subgroups of $\mathrm{P}B^\times$ of order 2: every nonscalar element $\alpha \in B^\times$ with trace zero has $\alpha^2 \in F^\times$.

Proposition 32.5.1. *Let $m > 2$ and let $\zeta_m \in F^{\mathrm{al}}$ be a primitive mth root of unity. Then $\mathrm{P}B^\times$ contains a cyclic subgroup of order m if and only if $\zeta_m + \zeta_m^{-1} \in F$ and $F(\zeta_m)$ splits B. Such a cyclic subgroup is unique up to conjugation in $\mathrm{P}B^\times$.*

Proof. First we prove ($\Leftarrow$). Suppose $\zeta_m + \zeta_m^{-1} \in F$ and $F(\zeta_m)$ splits B. If in fact $\zeta_m \in F$, then F splits B, i.e. $B \simeq \mathrm{M}_2(F)$; then $\gamma := \begin{pmatrix} 1 & 0 \\ 0 & \zeta_m \end{pmatrix}$ has order m in $\mathrm{P}B^\times \simeq \mathrm{PGL}_2(F)$. Otherwise, since $\zeta_m + \zeta_m^{-1} \in F$, we have $[F(\zeta_m) : F] = 2$, with ζ_m a root of the polynomial $T^2 - (\zeta_m + \zeta_m^{-1})T + 1$. By Lemma 5.4.7, this implies there is an embedding $F(\zeta_m) \hookrightarrow B$; let ζ be the image of ζ_m under this embedding. If ζ has order d in $\mathrm{P}B^\times$, then $\mathbb{Q}(\zeta_m + \zeta_m^{-1}, \zeta_m^d) \subseteq F$; since $\zeta_m \notin F$, we must have $d = m$ if m is odd or $d = m/2$ if m is even. Let $\gamma = 1 + \zeta$. Then $\gamma^2\zeta^{-1} = 2 + \zeta + \zeta^{-1} \in F^\times$, so γ has order m in $\mathrm{P}B^\times$.

Now we prove (⇒). Suppose that $\gamma \in B^\times$ has image in $\mathrm{P}B^\times$ of order $m > 2$, so that $\gamma^m = a \in F^\times$. We do calculations in the commutative F-algebra $K := F[\gamma]$. Let $\varsigma := \overline{\gamma}\gamma^{-1} \in K^\times$. Then

$$\varsigma^m = \overline{\gamma^m}(\gamma^m)^{-1} = aa^{-1} = 1 \tag{32.5.2}$$

so $\varsigma^m = 1$. If $\varsigma^d = 1$ for $d \mid m$ then $\overline{\gamma}^d = \varsigma^d\gamma^d = \gamma^d$ so $\gamma^d \in F^\times$ and thus $d = m$; thus ς has order m in $B^\times$. Applying the standard involution again gives

$$\gamma = \overline{\gamma\varsigma} = \overline{\varsigma}\overline{\gamma} = \varsigma\overline{\varsigma}\gamma; \tag{32.5.3}$$

thus $\overline{\varsigma} = \varsigma^{-1}$, so $\varsigma \notin F$ and $\operatorname{trd}(\varsigma) = \varsigma + \varsigma^{-1} \in F$. Taking an appropriate power to match up the root of unity, we conclude $\zeta_m + \zeta_m^{-1} \in F$. Finally, either K is a quadratic field in B, in which case K splits B by Lemma 5.4.7, or K is not a field and $B \simeq \mathrm{M}_2(F)$, in which case F already splits B.

We conclude with uniqueness. Continuing from the previous paragraph, we have shown that $\gamma + \overline{\gamma} = (1 + \varsigma)\gamma \in F^\times$, so γ and $1 + \varsigma$ generate the same cyclic subgroup of $\mathrm{P}B^\times$, where $\varsigma^m = 1$. If $K = F(\varsigma)$ is a field, then all embeddings $F(\zeta_m) \hookrightarrow B$ are conjugate in $B^\times$ by the Skolem–Noether theorem (Corollary 7.1.5), and consequently every two cyclic subgroups of order m are conjugate. Otherwise, the reduced characteristic polynomial of ς factors, so $B \simeq \mathrm{M}_2(F)$, and its roots (the eigenvalues of ς) belong to F. If the eigenvalues are repeated, then up to conjugation in $\mathrm{GL}_2(F)$, ς is a scalar multiple of $\begin{pmatrix} 1 & b \\ 0 & 1 \end{pmatrix}$ with $b \in F$, and therefore has infinite order, impossible. Thus the roots are distinct, and ς is conjugate to a multiple of $\begin{pmatrix} 1 & 0 \\ 0 & \lambda \end{pmatrix}$ and so λ is a primitive mth root of unity and the cyclic subgroup is unique up to conjugation. □

32.5.4. The splitting condition in Proposition 32.5.1 can alternatively be phrased in local-to-global terms (Proposition 14.6.7): $K = F(\zeta_m)$ splits B if and only if every place $v \in \operatorname{Ram} B$ is *not* split in K. Since the field $F(\zeta_m)$ is totally complex, every archimedean place splits, and so when $K \neq F$ we have $K \hookrightarrow B$ if and only if *no* prime $\mathfrak{p} \in \operatorname{Ram} B$ splits in K.

32.5.5. The proof of Proposition 32.5.1 describes the cyclic subgroup explicitly, up to conjugation (still with $m > 2$):

(i) If $\zeta_m \in F$, then $B \simeq \mathrm{M}_2(F)$ and every cyclic subgroup of $\mathrm{PGL}_2(F)$ of order m is conjugate to the subgroup generated by $\gamma_m = \begin{pmatrix} 1 & 0 \\ 0 & \zeta_m \end{pmatrix}$;

(ii) Otherwise, $K = F(\zeta_m)$ is a quadratic extension of F with $K \hookrightarrow B$, and every subgroup of $\mathrm{P}B^\times$ of order m is conjugate to the subgroup generated by the image of $\gamma_m = 1 + \zeta_m$.

The F-algebra $K_m = F[\gamma_m]$ is separable and uniquely determined up to isomorphism.

In contrast to Proposition 32.5.1, there are a great many cyclic subgroups of order $m = 2$ in $\mathrm{P}B^\times$, described as follows.

32.5.6. If $\gamma \in \mathrm{P}B^\times$ has order $m = 2$, then $\gamma^2 = a \in F^\times$ and $\gamma \notin F^\times$. Therefore, either $a \notin F^{\times 2}$, equivalently $K = F[\gamma] \simeq F(\sqrt{a})$ is a field, and the embedding $K \hookrightarrow B$ is unique up to conjugation in $B^\times$ by the Skolem–Noether theorem; or $a \in F^{\times 2}$, in which case after rescaling $\gamma^2 = 1$ so $B \simeq \mathrm{M}_2(F)$ and γ is conjugate to $\begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}$.

The following corollary shows that we can often reduce to the case of an even order subgroup.

Corollary 32.5.7. *Let $m \geq 1$ be odd. Then $\mathrm{P}B^\times$ contains a cyclic subgroup of order m if and only if $\mathrm{P}B^\times$ contains a cyclic subgroup of order $2m$.*

Proof. If $m = 1$, then we are all set by 32.5.6. If $m \geq 3$, then Proposition 32.5.1 applies, and we see that the hypotheses hold for m if and only if they hold for $2m$, since $\zeta_{2m} = -\zeta_m$. □

Corollary 32.5.8. *$\mathrm{P}B^\times$ contains a cyclic subgroup of order $2m$ if and only if $\mathrm{P}B^1$ contains a cyclic subgroup of order m.*

Proof. The corollary follows from 32.5.5: the subgroup of $\mathrm{P}B^\times$ of order $2m$ generated by γ_{2m} yields the subgroup of $\mathrm{P}B^1$ of order m generated by ζ_{2m}, and vice versa. □

32.6 ∗ Dihedral subgroups

We now turn to the dihedral case, where we show that every cyclic subgroup extends (in general, in many ways) to a dihedral subgroup, continuing to follow Chinburg–Friedman [CF2000, Lemma 2.3].

Lemma 32.6.1. *Let $m \geq 2$. Then the following statements hold.*

(a) *Every cyclic subgroup of $\mathrm{P}B^\times$ of order m is contained in a dihedral subgroup of order $2m$; in particular, $\mathrm{P}B^\times$ contains a dihedral subgroup of order $2m$ if and only if it contains a cyclic subgroup of order m.*
(b) *Let $\gamma \in B^\times$ have order m in $\mathrm{P}B^\times$, and let $K = F[\gamma]$. For $j \in B^\times$, we have $\langle \gamma, j \rangle \subseteq \mathrm{P}B^\times$ dihedral if and only if $j^2 = b \in F$ and $B = K + Kj \simeq \left(\dfrac{K, b}{F}\right)$.*

Proof. First (a). The implication ($\Rightarrow$) is immediate, so we prove ($\Leftarrow$). Let $m \geq 2$ and suppose that $\mathrm{P}B^\times$ contains a cyclic subgroup of order m, generated by the image of $\gamma \in B^\times$, and let $K = F[\gamma]$. Let $j \in B$ be orthogonal to K under nrd. Then $j^2 = b \in F^\times$, and $j\alpha = \overline{\alpha} j = \alpha^{-1} j \in \mathrm{P}B^\times$ for all $\alpha \in K$, so the subgroup $\langle \gamma, j \rangle$ is dihedral of order $2m$ in $\mathrm{P}B^\times$, and $B \simeq \left(\dfrac{K, b}{F}\right)$, as in Exercise 6.2.

Now (b). We just showed ($\Rightarrow$) in the previous part, so we show ($\Leftarrow$). Let $\Gamma \subseteq \mathrm{P}B^\times$ be a dihedral subgroup of order $2m$, where $\Gamma = \langle \gamma, j \rangle$ has $\gamma \in B^\times$ generating a cyclic subgroup of order m in $\mathrm{P}B^\times$ and $j \in B^\times$ satisfies

$$j^{-1} \gamma j = \gamma^{-1} \in B^\times / F^\times.$$

Let $K = F[\gamma]$. We claim that $B = K + Kj \simeq \left(\dfrac{K, b}{F}\right)$.

First we show $j^{-1}\gamma j = \overline{\gamma}$. This follows from a direct argument using reduced norm and trace (see Exercise 32.4), but we have also the following argument. Since $K = F[\gamma]$ is semisimple (see 32.5.5) and conjugation by j acts as an F-algebra automorphism of $K = F[\gamma]$, it is either the identity or the standard involution, and thus $j^{-1}\gamma j = \gamma, \overline{\gamma}$. But we cannot have $j^{-1}\gamma j = \gamma$, because then $K[j] \subseteq B$ would be a commutative subalgebra of dimension ≥ 3, a contradiction.

Now by (4.2.16), expanding the trace gives

$$\begin{aligned} \operatorname{trd}(j\overline{\gamma}) &= j\overline{\gamma} + \gamma\overline{j} = j\overline{\gamma} + \gamma(\operatorname{trd}(j) - j) \\ &= j\overline{\gamma} - \gamma j + \operatorname{trd}(j)\gamma = \operatorname{trd}(j)\gamma. \end{aligned} \tag{32.6.2}$$

Since $1, \gamma$ are linearly independent we conclude $\operatorname{trd}(j) = \operatorname{trd}(j\overline{\gamma}) = 0$, i.e., j is orthogonal to K under nrd, so $j^2 = b \in F^\times$ and $B = \left(\dfrac{K, b}{F}\right)$. ◻

The dihedral subgroups of order $2m$ for $m > 2$ are classified as follows.

Lemma 32.6.3. *Let $m > 2$. Then the set of dihedral subgroups of order $2m$ up to conjugation in $\mathrm{P}B^\times$ are in bijection with the group*

$$\frac{\operatorname{Nm}_{K_m/F}(K_m^\times)}{\langle\delta\rangle F^{\times 2}} \tag{32.6.4}$$

where K_m is as in 32.5.5 and $\delta = 2 + \zeta_m + \zeta_m^{-1}$.

Proof. Let $\Gamma = \langle \gamma, j\rangle$ and $\Gamma' = \langle \gamma, j'\rangle$ be two dihedral subgroups as in Lemma 32.6.1(b) with $j^2 = b$ and $(j')^2 = b'$. Then $j' \in K^\perp = Kj$ so $j' = \beta j$ with $\beta \in K^\times$, and $\operatorname{nrd}(j') = b' = \operatorname{nrd}(\beta)b$, so $bb' \in \operatorname{Nm}_{K|F}(K^\times)$ (as in Exercise 6.4). Then we claim that Γ, Γ' are conjugate in $\mathrm{P}B^\times$ if and only if

$$bb' \in F^{\times 2}\langle\delta\rangle.$$

If $\Gamma' = \alpha^{-1}\Gamma\alpha$ with $\alpha \in B^\times$, then conjugation by α normalizes the unique cyclic subgroup on both sides, so $\gamma^r j' = \alpha^{-1} j\alpha$ for some r, and therefore

$$-\delta^r b' = \operatorname{nrd}(\gamma^r j') = \operatorname{nrd}(j) = -b$$

as desired. Conversely, if $bb' \in F^{\times 2}\langle\delta\rangle$ then so too for $\operatorname{nrd}(\beta) = b'b^{-1}$; rescaling β by $F^\times$ and replacing β by $\gamma\beta$ if necessary, we may suppose $\operatorname{nrd}(\beta) = \operatorname{Nm}_{K|F}(\beta) = 1$ (without changing Γ'); by Hilbert's theorem 90, there exists $\alpha \in K^\times$ such that $\beta = \overline{\alpha}\alpha^{-1}$, and conjugation by α again normalizes the cyclic subgroup and satisfies

$$\alpha^{-1} j\alpha = \overline{\alpha}\alpha^{-1} j = \beta j$$

as desired. ◻

Remark 32.6.5. One can rephrase Lemma 32.6.3 in terms of a global equivalence relation, further encompassing the case $m = 2$: see Chinburg–Friedman [CF2000, Lemma 2.4].

Corollary 32.6.6. *The group* $\mathrm{P}B^1$ *contains a dihedral group of order* $2m > 4$ *if and only if* $B \simeq \left(\dfrac{K_{2m}, -1}{F}\right)$.

Proof. We first prove ($\Rightarrow$). If $\mathrm{P}B^1$ contains a dihedral group $\Gamma = \langle \gamma, j \rangle$ of order $2m$ then it contains a cyclic subgroup of order m so by Corollary 32.5.8 the group $\mathrm{P}B^\times$ contains a cyclic subgroup of order $2m$, which we may take to be generated by γ_{2m} as in 32.5.5 with $K_{2m} = F[\gamma_{2m}] = F[\gamma]$; by hypothesis we have $j^2 = -\operatorname{nrd}(j) = -1$, and so $B \simeq \left(\dfrac{K_{2m}, -1}{F}\right)$ as in the classification in Lemma 32.6.1.

Next we prove the converse implication ($\Leftarrow$). We refer to 32.5.5. In case (i) where $\zeta_{2m} \in F$, we have $\operatorname{nrd}(\zeta_{2m}^{-1}\gamma_m) = \zeta_{2m}^2\zeta_m = 1$ so we may take $\Gamma = \langle \zeta_{2m}^{-1}\gamma_m, j \rangle$; in case (ii), where $\zeta_{2m} \notin F$, we take $\Gamma = \langle \zeta_{2m}, j \rangle$. ◻

32.7 ∗ Exceptional subgroups

Finally we treat exceptional groups (cf. Gehring–Maclachlan–Martin–Reid [GMMR97, p. 3635]). We found quaternionic realizations of the exceptional groups in section 11.5 when $B \simeq (-1, -1 \mid F)$ (and $\sqrt{5} \in F$ for A_5).

Proposition 32.7.1. *The following statements hold.*

(a) $\mathrm{P}B^\times$ *contains a subgroup isomorphic to* A_4 *if and only if* $\mathrm{P}B^1$ *contains a subgroup isomorphic to* A_4 *if and only if* $B \simeq (-1, -1 \mid F)$.
(b) $\mathrm{P}B^\times$ *contains a subgroup isomorphic to* S_4 *if and only if it contains a subgroup isomorphic to* A_4*; and* $\mathrm{P}B^1$ *contains a subgroup isomorphic to* S_4 *if and only if* $B \simeq (-1, -1 \mid F)$ *and* $\sqrt{2} \in F$.
(c) $\mathrm{P}B^\times$ *contains a subgroup isomorphic to* A_5 *if and only if* $\mathrm{P}B^1$ *contains a subgroup isomorphic to* A_5 *if and only if* $B \simeq (-1, -1 \mid F)$ *and* $\sqrt{5} \in F$.

Any two such exceptional subgroups of $\mathrm{P}B^\times$ *(or* $\mathrm{P}B^1$*) are conjugate by an element of* $B^\times$ *if and only if they are isomorphic as groups.*

Proof. First we prove (a); let $\Gamma \subseteq \mathrm{P}B^\times$ be a subgroup with $\Gamma \simeq A_4$. The reduced norm gives a homomorphism $\Gamma \to \operatorname{nrd}(\Gamma) \subseteq F^\times/F^{\times 2}$, but A_4 has no nontrivial homomorphic image of exponent 2, so $\operatorname{nrd}(\Gamma) \subseteq F^{\times 2}$. Therefore, there is a unique lift of Γ to $B^1/\{\pm 1\}$, and the map $B^1/\{\pm 1\} \to B^\times/F^\times$ is an isomorphism from this lift to H. This shows the first implication; its converse follows from the injection $\mathrm{P}B^1 \hookrightarrow \mathrm{P}B^\times$. For the second implication, let $i, j \in B^1$ generate the V_4-subgroup (the normal subgroup of index 3 isomorphic to the Klein 4 group) of A_4 in $\mathrm{P}B^1$. Then $i, j \notin F^\times$, and $i^2 = -\operatorname{nrd}(i) = -1 = j^2$; and similarly $(ij)^2 = -1$ implies $ji = -ij$. By Lemma 2.2.5, we conclude $B \simeq (-1, -1 \mid F)$. The converse follows from the Hurwitz unit group 11.2.4.

For part (b), the implication (⇒) is immediate; the implication (⇐) follows by taking the Hurwitz units and adjoining the element $1+i$, as in 11.5.4 but working modulo scalars. For the second statement, an element of order 4 in $B^1/\{\pm 1\}$ lifts to an element of order 8 in B^1 and therefore has reduced trace $\pm\sqrt{2} \in F$; the converse follows again from the explicit construction in 11.5.4.

For part (c), we argue similarly. Since A_5 is generated by its subgroups isomorphic to A_4, we may apply (a) to get a lift, and by the reduced trace we get $\sqrt{5} \in F$.

The uniqueness statement is requested in Exercise 32.5. □

Exercises

Unless otherwise indicated, let F be a number field with ring of integers R, let B be a quaternion algebra over F and let $O \subseteq B$ be an R-order.

1. Let F be a totally real field, let $K \supseteq F$ be a totally imaginary quadratic extension of F, so $K = F(\sqrt{d})$ with d totally negative. Let S be the ring of integers of K. Consider the group homomorphism

$$\begin{aligned} \phi\colon S^\times &\to S^\times \\ u &\mapsto u/\overline{u} \end{aligned}$$

where $\overline{}$ is the nontrivial F-involution of K.

   (a) Show that if $u \in \mu(S)R^\times$ then $u \in \ker\phi$.
   (b) Show that $\phi(u)$ is a root of unity for all $u \in S^\times$. *[Hint: It is an algebraic integer of absolute value* 1 *under all complex embeddings.]*
   (c) Let $\mu(S)$ be the subgroup of roots of unity of $S^\times$, and let $\psi\colon S^\times \to \mu(S)/\mu(S)^2$ be the map induced by ϕ. Show that if $u \in \ker\psi$, so $\phi(u) = \zeta^2$ with $\zeta \in \mu(S)$, then $\zeta^{-1}u \in R^\times$. Conclude that $\ker\psi = \mu(S)R^\times$.
   (d) Show that $[S^\times : \mu(S)R^\times] \leq [S^\times : S^1R^\times] \leq 2$.

   [The index $[S^\times : \mu(S)R^\times]$ is known as the **Hasse unit index**.]

2. Let $B = \left(\dfrac{-1,-1}{\mathbb{Q}}\right)$ and let O be the $\mathbb{Z}$-order generated by i, j. Prove that $O[1/2]^\times \simeq \langle 2, i, j, 1+i\rangle$ and describe $O[1/2]^\times/\mathbb{Z}[1/2]^\times$ as an extension of Q_8 by $\mathbb{Z}/2\mathbb{Z}$.

3. Show that

$$[O^\times : O^1R^\times] = [\operatorname{nrd}(O^\times) : R^{\times 2}] = \frac{[R^\times_\Omega : R^{\times 2}]}{[R^\times_\Omega : \operatorname{nrd}(O^\times)]}$$

where $R^\times_{>_\Omega 0}$ is the subgroup of units positive at ramified infinite places of B, defined in (32.2.6).

4. Let B be a quaternion algebra over a field F with $\operatorname{char} F \neq 2$, and let $\gamma, j \in B^\times$ be such that $j^{-1}\gamma j = \gamma^{-1}$. Show by looking at the reduced norm and trace that $j^{-1}\gamma j = \overline{\gamma}$ (cf. Lemma 32.6.1).

▹ 5. Prove the uniqueness statement in Proposition 32.7.1: Show that every two isomorphic exceptional subgroups of $\mathrm{P}B^\times$ are conjugate by an element of $B^\times$, and the same for B^1.

6. Let $\Gamma \leq \mathcal{O}^1$ be a maximal finite subgroup. Combining results from sections 11.2 and 11.5 and this chapter, prove the following.

   (a) Γ is isomorphic to one of the following groups:
   - cyclic of order $2m$ with $m \geq 1$,
   - binary dihedral (dicyclic) $2D_{2m}$ of order $4m$ with $m \geq 1$,
   - binary tetrahedral $2T$ of order 24,
   - binary octahedral $2O$ of order 48, or
   - binary icosahedral $2I$ of order 120.

   In the latter three cases, we call Γ **exceptional**.

   (b) If $\Gamma \simeq 2O$ then $F \supseteq \mathbb{Q}(\sqrt{2})$ and if $\Gamma \simeq 2I$ then $F \supseteq \mathbb{Q}(\sqrt{5})$.

   (c) If Γ is exceptional, then $B \simeq \left(\dfrac{-1,-1}{F}\right)$.

7. Continuing with the previous exercise, suppose that B is totally definite, so F is totally real. Prove the following statements.

   (a) If $\mathcal{O}^1$ does not contain an element of order 4, then $\mathcal{O}^1$ is cyclic.
   (b) If $\mathcal{O}^1$ is quaternion $Q_8 \simeq 2D_4$ or exceptional, then $R[\sqrt{-1}] \hookrightarrow \mathcal{O}$ and $\operatorname{discrd}(\mathcal{O})$ is only divisible by primes dividing 2.
   (c) If $\mathcal{O}^1 \simeq 2D_{2m}$ with $m \geq 3$, then $R[\sqrt{-1}], R[\zeta_{2m}] \hookrightarrow \mathcal{O}$ and $\operatorname{discrd}(\mathcal{O})$ is only divisible by primes dividing $\lambda_{2m}^2 - 4$, where $\lambda_{2m} := \zeta_{2m} + \zeta_{2m}^{-1}$.

8. Continuing further with the previous exercise, we compare $\mathcal{O}^\times$ and $\mathcal{O}^1$.

   (a) Show that $[\mathcal{O}^\times : \mathcal{O}^1] = 1, 2, 4$. *[Hint: Use Exercise 32.1.]*
   (b) Show that if $\mathcal{O}^1 \simeq 2O, 2I$, then $\mathcal{O}^\times = \mathcal{O}^1$.
   (c) Show that if $\mathcal{O}^1 \simeq 2T$, then $[\mathcal{O}^\times : \mathcal{O}^1] \leq 2$, and equality holds if and only if there exists $\gamma \in (1+i)F^\times \cap \mathcal{O}^\times$ such that $\operatorname{nrd}(\gamma) \notin F^{\times 2}$.

   [For a complete account covering all cases, see Vignéras-Guého [VG74].]

This chapter is licensed under the terms of the Creative Commons Attribution-NonCommercial 4.0 International License (http://creativecommons.org/licenses/by-nc/4.0/), which permits any noncommercial use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license and indicate if changes were made.

The images or other third party material in this chapter are included in the chapter's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the chapter's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.

Chapter 33

Hyperbolic plane

In this chapter, we give background on the geometry of the hyperbolic plane.

33.1 ▹ The beginnings of hyperbolic geometry

We have seen that the group of unit Hamiltonians $\mathbb{H}^1$ acts by rotations of Euclidean space and therefore by isometries of the unit sphere, and that in spherical geometry the discrete subgroups are beautifully realized as classical finite groups: cyclic, dihedral, and the symmetry groups of the Platonic solids.

Replacing $\mathbb{H}$ with $\mathrm{M}_2(\mathbb{R})$, the group $\mathrm{SL}_2(\mathbb{R})$ of determinant 1 matrices possesses a much richer supply of discrete subgroups. In fact, $\mathrm{PSL}_2(\mathbb{R})$ can be naturally identified with a circle bundle over the hyperbolic plane, and so the structure of quaternionic unit groups is naturally phrased in the language of hyperbolic geometry. Indeed, it was work on automorphic functions and differential equations invariant under discrete subgroups of $\mathrm{PSL}_2(\mathbb{R})$ that provided additional early original motivation to study hyperbolic space: their study provides an incredibly rich interplay between number theory, algebra, geometry, and topology, with quaternionic applications. This interplay is the subject of the final parts of the text.

In this chapter, we provide a rapid introduction to the hyperbolic plane. Hyperbolic geometry has its roots preceding the quaternions, in efforts during the early 1800s to understand Euclid's axioms for geometry. Since the time of Euclid, there had been attempts to prove the quite puzzling parallel postulate (given a line and a point not on the line, there is a unique line through the point parallel to the given line) from the other four simpler, self-evident axioms for geometry. In hyperbolic geometry, the parallel postulate fails to hold—there are always infinitely many distinct lines through a point that do not intersect a given line—and so it is a non-Euclidean geometry.

The underpinnings of what became hyperbolic geometry can be found in work by Euler and Gauss in their study of curved surfaces (the differential geometry of surfaces). It was then Lobachevsky and Bolyai who suggested that curved surfaces of constant negative curvature could be used in non-Euclidean geometry, and finally Riemann who generalized this to what are now called Riemannian manifolds. Klein coined the term "hyperbolic" for this geometry because its formulae can be obtained from spherical

© The Author(s) 2021

J. Voight, *Quaternion Algebras*, Graduate Texts in Mathematics 288,
https://doi.org/10.1007/978-3-030-56694-4_33

geometry by replacing trigonometric functions by their hyperbolic counterparts. See [Sco83, §1] for a nice overview of the 2-dimensional geometries.

Hyperbolic geometry, and in particular the hyperbolic plane, remains an important prototype for understanding negatively-curved spaces in general. Milnor [Milno82] gives a comprehensive early history of hyperbolic geometry; see also the survey by Cannon–Floyd–Kenyon–Parry [CFKP97], which includes an exposition of five models for hyperbolic geometry. (It is also possible to work out hyperbolic geometry in a manner akin to what Euclid did for his geometry without a particular model, following Lobachevsky [LP2010].)

For further references on hyperbolic plane geometry, see Jones–Singerman [JS87, Chapter 5], Anderson [And2005], Ford [For72], Katok [Kat92, Chapter 1], Iversen [Ive92, Chapter III], and Beardon [Bea95, Chapter 7]. There are a wealth of geometric results and formulas from Euclidean geometry that one can try to reformulate in the world of hyperbolic plane geometry, and the interested reader is encouraged to pursue these further.

33.2 Geodesic spaces

In geometry, we need notions of length, distance, and the straightness of a path. These notions are defined for a certain kind of metric space, as follows.

Let X be a metric space with distance ρ. An **isometry** $g : X \xrightarrow{\sim} X$ is a bijective map that preserves distance, i.e., $\rho(x, y) = \rho(g(x), g(y))$ for all $x, y \in X$. (Any distance-preserving map is automatically injective and so becomes an isometry onto its image.) The set of isometries $\mathrm{Isom}(X)$ forms a group under composition.

33.2.1. A **path** from x to y, denoted $\upsilon\colon x \to y$, is a continuous map $\upsilon\colon [0,1] \to X$ where $\upsilon(0) = x$ and $\upsilon(1) = y$. (More generally, we can take the domain to be any compact real interval.) The **length** $\ell(\upsilon)$ of a path υ is the supremum of sums of distances between successive points over all finite subdivisions of the path (the path is **rectifiable** if this supremum is finite). Conversely, if X is a set with a notion of (nonnegative) length of path, then one recovers a candidate (**intrinsic**) metric as

$$\rho(x, y) = \inf_{\upsilon : x \to y} \ell(\upsilon), \tag{33.2.2}$$

a metric when this infimum exists (i.e., there exists a path of finite length $x \to y$) for all $x, y \in X$. If the distance on X is of the form (33.2.2), we call X a **length metric space** or a **path metric space**, and by construction $\ell(g\upsilon) = \ell(\upsilon)$ for all paths υ and $g \in \mathrm{Isom}(X)$.

Example 33.2.3. The space $X = \mathbb{R}^n$ with the ordinary Euclidean metric is a path metric space; it is sometimes denoted E^n as Euclidean space (to emphasize the role of the metric).

33.2.4. If X is a path metric space and υ achieves the infimum in (33.2.2), then we say υ is a **geodesic segment** in X. A **geodesic** is a continuous map $(-\infty, \infty) \to X$ such that the restriction to every compact interval defines a geodesic segment. If X is a path

metric space such that every two points in X are joined by a geodesic segment, we say X is a **geodesic space**, and if this geodesic is unique we call X a **uniquely geodesic space**.

33.2.5. If X is a geodesic space, then an isometry of X maps geodesic segments to geodesic segments, and hence geodesics to geodesics: i.e., if $g \in \mathrm{Isom}(X)$ and $\upsilon\colon x \to y$ is a geodesic segment, then $g\upsilon\colon gx \to gy$ is a geodesic segment. After all, g maps the set of paths $x \to y$ bijectively to the set of paths $gx \to gy$, preserving distance.

33.2.6. In the context of differential geometry (our primary concern), these notions can be made concrete with coordinates. Suppose $U \subseteq \mathbb{R}^n$ is an open subset. Then a convenient way to specify the length of a path in U is with a **length element** in real-valued coordinates. To illustrate, the ordinary metric on $\mathbb{R}^n$ is given by the length element

$$\mathrm{d}s := \sqrt{\mathrm{d}x_1^2 + \cdots + \mathrm{d}x_n^2},$$

so if $\upsilon\colon [0,1] \to U$ is a piecewise continuously differentiable function written as $\upsilon(t) = (x_1(t), \ldots, x_n(t))$, then

$$\ell(\upsilon) = \int_0^1 \sqrt{\left(\frac{\mathrm{d}x_1}{\mathrm{d}t}\right)^2 + \cdots + \left(\frac{\mathrm{d}x_n}{\mathrm{d}t}\right)^2}\,\mathrm{d}t \tag{33.2.7}$$

as usual.

More generally, if $\lambda\colon U \to \mathbb{R}_{>0}$ is a positive continuous function, then the length element $\lambda(x)\,\mathrm{d}s$ defines a metric (33.2.2) on U, as follows. The associated length (33.2.7) is symmetric, nonnegative, and satisfies the triangle inequality. To show that $\rho(x, y) > 0$ when $x \neq y$, by continuity λ is bounded below by some $\eta > 0$ on a suitably small ϵ ball neighborhood of x not containing y, so every path $\upsilon\colon x \to y$ has $\ell(\upsilon) \geq \epsilon\eta$ and $\rho(x, y) > 0$.

In this context, we also have a notion of orientation, and we may restrict to isometries that preserve this orientation. We return to this in section 33.8, rephrasing this in terms of Riemannian geometry.

Remark 33.2.8. The more general study of geometry based on the notion of length in a topological space (the very beginnings of which are presented here) is the area of *metric geometry*. Metric geometry has seen significant recent applications in group theory and dynamical systems, as well as many other areas of mathematics. For further reading, see the texts by Burago–Burago–Ivanov [BBI2001] and Papadopoulous [Pap2014].

In particular, geodesic spaces are quite common in mathematics, including complete Riemannian manifolds; Busemann devotes an entire book to the geometry of geodesics [Bus55]. Uniquely geodesic spaces are less common; examples include simply connected Riemannian manifolds without conjugate points, CAT(0) spaces, and Busemann convex spaces.

The following theorem nearly characterizes geodesic spaces.

Theorem 33.2.9 (Hopf–Rinow). *Let X be a complete and locally compact length metric space. Then X is a geodesic space and every bounded closed set in X is compact.*

Proof. See e.g. Bridson–Haefliger [BH99, Proposition 3.7]). □

33.3 Upper half-plane

We now present the first model of two-dimensional hyperbolic space (see Figure 33.3.2).

Definition 33.3.1. The **upper half-plane** is the set

$$\mathbf{H}^2 := \{z = x + iy \in \mathbb{C} : \operatorname{Im}(z) = y > 0\}.$$

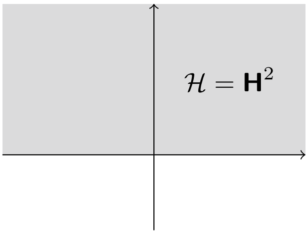


Figure 33.3.2: Upper half-plane $\mathbf{H}^2$

Definition 33.3.3. The **hyperbolic length element** on $\mathbf{H}^2$ is defined by

$$\mathrm{d}s := \frac{\sqrt{\mathrm{d}x^2 + \mathrm{d}y^2}}{y} = \frac{|\mathrm{d}z|}{\operatorname{Im} z}; \tag{33.3.4}$$

As described in 33.2.6, the hyperbolic length element induces a metric on $\mathbf{H}^2$, and this provides it with the structure of a path metric space.

Definition 33.3.5. The set $\mathbf{H}^2$ equipped with the hyperbolic metric is (a model for) the **hyperbolic plane**.

Remark 33.3.6. The space $\mathbf{H}^2$ can be intrinsically characterized as the unique two-dimensional (connected and) simply connected Riemannian manifold with constant sectional curvature -1.

The hyperbolic metric and the Euclidean metric on $\mathbf{H}^2$ are equivalent, inducing the same topology (Exercise 33.1). However, lengths and geodesics are different under these two metrics, as we will soon see.

33.3.7. The group $\mathrm{GL}_2(\mathbb{R})$ acts on $\mathbb{C}$ via linear fractional transformations:

$$gz = \frac{az+b}{cz+d}, \qquad \text{for } g = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in \mathrm{SL}_2(\mathbb{R}) \text{ and } z \in \mathbb{C};$$

since

$$gz = \frac{(az+b)(\overline{cz+d})}{|cz+d|^2} = \frac{ac|z|^2 + adz + bc\overline{z} + bd}{|cz+d|^2} \tag{33.3.8}$$

we have

$$\operatorname{Im} gz = \frac{\det g}{|cz+d|^2} \operatorname{Im} z. \tag{33.3.9}$$

and so if $\operatorname{Im} z > 0$, then $\operatorname{Im} gz > 0$ if and only if $\det g > 0$. Therefore, the subgroup

$$\mathrm{GL}_2^+(\mathbb{R}) = \{g \in \mathrm{GL}_2(\mathbb{R}) : \det(g) > 0\}$$

preserves the upper half-plane $\mathbf{H}^2$. Moreover, because the action of $\mathrm{GL}_2^+(\mathbb{R})$ is holomorphic, it is orientation-preserving.

The kernel of this action, those matrices acting by the identity as linear fractional transformations, are the scalar matrices, since $(az+b)/(cz+d) = z$ identically if and only if $c = b = 0$ and $a = d$. Taking the quotient we get a faithful action of $\mathrm{PGL}_2^+(\mathbb{R}) = \mathrm{GL}_2^+(\mathbb{R})/\mathbb{R}^\times$ on $\mathbf{H}^2$. There is a canonical isomorphism

$$\begin{aligned} \mathrm{PGL}_2^+(\mathbb{R}) &\xrightarrow{\sim} \mathrm{PSL}_2(\mathbb{R}) \\ g &\mapsto \frac{1}{\sqrt{\det(g)}} g. \end{aligned}$$

with the same action on the upper half-plane.

33.3.10. The determinant $\det : \mathrm{PGL}_2(\mathbb{R}) \to \mathbb{R}^\times/\mathbb{R}^{\times 2} \simeq \{\pm 1\}$ has the inverse image of $+1$ equal to $\mathrm{PGL}_2^+(\mathbb{R})$ both open and closed in $\mathrm{PGL}_2(\mathbb{R})$; therefore, any g with $\det(g) < 0$ together with $\mathrm{PGL}_2^+(\mathbb{R})$ generates $\mathrm{PGL}_2(\mathbb{R})$: for example, we may take

$$g = \begin{pmatrix} -1 & 0 \\ 0 & 1 \end{pmatrix} \tag{33.3.11}$$

In view of (33.3.9), we extend the action of $\mathrm{PGL}_2(\mathbb{R})$ on $\mathbf{H}^2$ by defining for $g \in \mathrm{PGL}_2(\mathbb{R})$ and $z \in \mathbf{H}^2$

$$g \cdot z = \begin{cases} gz, & \text{if } \det g > 0; \\ g\overline{z} = \overline{gz}, & \text{if } \det g < 0. \end{cases} \tag{33.3.12}$$

The elements $g \in \mathrm{PGL}_2(\mathbb{R})$ with $\det g < 0$ act anti-holomorphically and so are orientation-reversing. The matrix g in (33.3.11) then acts by $g(z) = -\overline{z}$.

This action also arises naturally from another point of view. Let

$$\mathbf{H}^{2-} = \{z \in \mathbb{C} : \operatorname{Im} z < 0\}$$

be the lower half-plane, let $\mathbf{H}^{2+} = \mathbf{H}^2$, and let

$$\mathbf{H}^{2\pm} = \mathbf{H}^{2+} \cup \mathbf{H}^{2-} = \{z \in \mathbb{C} : \operatorname{Im} z \neq 0\} = \mathbb{C} - \mathbb{R}.$$

Then $\mathrm{PGL}_2(\mathbb{R})$ acts on $\mathbf{H}^{2\pm}$ (it preserves $\mathbb{R}$ hence also its complement in $\mathbb{C}$). Complex conjugation $z \mapsto \overline{z}$ interchanges $\mathbf{H}^{2+}$ and $\mathbf{H}^{2-}$, and there is a canonical identification

$$\mathbf{H}^{2\pm}/\langle \overline{} \rangle \xrightarrow{\sim} \mathbf{H}^2$$

from which we obtain the action (33.3.12) of $\mathrm{PGL}_2(\mathbb{R})$ on $\mathbf{H}^2$.

Remark 33.3.13. The fact that $\mathrm{PSL}_2(\mathbb{R})$ has elements $g \in \mathrm{PSL}_2(\mathbb{R})$ that are matrices up to sign means that whenever we do a computation with a choice of matrix, implicitly we are also checking that the computation goes through with the other choice of sign. Most of the time, this is harmless—but in certain situations this sign plays an important role!

Let $\mathrm{Isom}^+(\mathbf{H}^2) \leq \mathrm{Isom}(\mathbf{H}^2)$ be the subgroup of isometries of $\mathbf{H}^2$ that preserve orientation.

Theorem 33.3.14. *The group* $\mathrm{PSL}_2(\mathbb{R})$ *acts on* $\mathbf{H}^2$ *via orientation-preserving isometries, i.e.,* $\mathrm{PSL}_2(\mathbb{R}) \hookrightarrow \mathrm{Isom}^+(\mathbf{H}^2)$.

Proof. Because the metric is defined by a length element $\mathrm{d}s$, we want to show that $\mathrm{d}(gs) = \mathrm{d}s$ for all $g \in \mathrm{PSL}_2(\mathbb{R})$, i.e.,

$$\frac{|\mathrm{d}(gz)|}{\mathrm{Im}(gz)} = \frac{|\mathrm{d}z|}{\mathrm{Im}\,z}$$

for all $g \in \mathrm{PSL}_2(\mathbb{R})$. Since $|\mathrm{d}(gz)| = |\mathrm{d}g(z)/\mathrm{d}z||\mathrm{d}z|$, it is equivalent to show that

$$\frac{|\mathrm{d}g(z)/\mathrm{d}z|}{\mathrm{Im}\,gz} = \frac{1}{\mathrm{Im}\,z} \tag{33.3.15}$$

for all $g \in \mathrm{PSL}_2(\mathbb{R})$.

Let $g \in \mathrm{PSL}_2(\mathbb{R})$ act by

$$g(z) = \frac{az+b}{cz+d}$$

with $ad - bc = 1$. Then

$$\left|\frac{\mathrm{d}g}{\mathrm{d}z}(z)\right| = \left|\frac{(cz+d)a-(az+b)c}{(cz+d)^2}\right| = \frac{1}{|cz+d|^2}; \tag{33.3.16}$$

by (33.3.9),

$$\mathrm{Im}\,gz = \frac{\mathrm{Im}\,z}{|cz+d|^2},$$

so taking the ratio, the two factors $|cz+d|^2$ exactly cancel, establishing (33.3.15).

Finally, the action is holomorphic so (by the Cauchy–Riemann equations) it lands in the orientation-preserving subgroup. □

33.3.17. The action of $\mathrm{PSL}_2(\mathbb{R})$ extends to the boundary as follows. We define the **circle at infinity** to be the boundary

$$\mathrm{bd}\,\mathbf{H}^2 := \mathbb{R} \cup \{\infty\} \subseteq \mathbb{C} \cup \{\infty\}.$$

(The name comes from viewing $\mathbf{H}^2$ in stereographic projection as a half-sphere with circular boundary.) The group $\mathrm{PSL}_2(\mathbb{R})$ still acts on $\mathrm{bd}\,\mathbf{H}^2$ by linear fractional transformations. We define the **completed upper half-plane** to be

$$\mathbf{H}^{2*} := \mathbf{H}^2 \cup \mathrm{bd}\,\mathbf{H}^2.$$

The topology on $\mathbf{H}^{2*}$ is the same as the Euclidean topology on $\mathbf{H}^2$, and we take a fundamental system of neighborhoods of the point ∞ to be sets of the form

$$\{z \in \mathbf{H}^2 : \operatorname{Im} z > M\} \cup \{\infty\}$$

for $M > 0$ and a system of neighborhoods of the point z_0 to be

$$\{z_0\} \cup \{|z - (\operatorname{Re} z_0 + mi)| < m\}$$

i.e. open disks tangent to the real axis at z_0, together with z_0.

Remark 33.3.18. Although the hyperbolic plane cannot be embedded in $\mathbb{R}^3$, it can be *locally* embedded. One way to visualize plane hyperbolic geometry (locally) is by the **pseudosphere**, the surface of revolution generated by a tractrix: it is a surface with constant negative curvature and so locally models the hyperbolic plane (Figure 33.3.19).

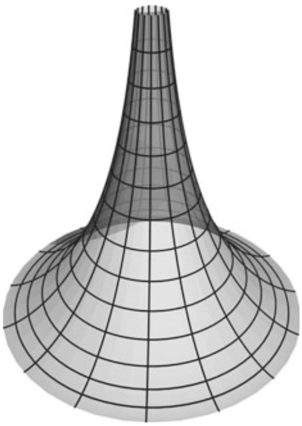


Figure 33.3.19: Pseudosphere

Remark 33.3.20. We will compactify quotients of $\mathbf{H}^2$ in other ways below. In that context, we will add only a subset of the boundary of $\mathbf{H}^{2*}$; this overloading should cause no confusion.

33.4 Classification of isometries

On our way to classifying isometries, we pause to identify three natural subgroups of $\mathrm{SL}_2(\mathbb{R})$:

$$\begin{aligned}
K = \mathrm{SO}(2) &= \left\{\begin{pmatrix} \cos t & \sin t \\ -\sin t & \cos t \end{pmatrix} : t \in \mathbb{R}\right\} \simeq \mathbb{R}/2\pi\mathbb{Z} \\
A &= \left\{\begin{pmatrix} a & 0 \\ 0 & 1/a \end{pmatrix} : a \in \mathbb{R}_{>0}^{\times}\right\} \simeq \mathbb{R}_{>0}^{\times} \simeq \mathbb{R} \qquad (33.4.1) \\
N &= \left\{\begin{pmatrix} 1 & b \\ 0 & 1 \end{pmatrix} : b \in \mathbb{R}\right\} \simeq \mathbb{R}.
\end{aligned}$$

We have $K = \mathrm{Stab}_{\mathrm{SL}_2(\mathbb{R})}(i)$ since $\dfrac{ai+b}{ci+d} = i$ if and only if $d = a$ and $c = -b$, and then the determinant condition implies $a^2 + b^2 = 1$. An element $\begin{pmatrix} a & 0 \\ 0 & 1/a \end{pmatrix}$ acts by $z \mapsto a^2 z$, fixing the origin and stretching along lines through the origin. An element $\begin{pmatrix} 1 & n \\ 0 & 1 \end{pmatrix}$ acts by the translation $z \mapsto z + b$.

Proposition 33.4.2 (Iwasawa decomposition). *The multiplication map gives a homeomorphism*

$$N \times A \times K \xrightarrow{\sim} \mathrm{SL}_2(\mathbb{R}).$$

In particular, for all $g \in \mathrm{SL}_2(\mathbb{R})$, we can write uniquely $g = n_g a_g k_g$ with $n_g \in N$, $a_g \in A$, and $k_g \in K$ in a way continuously varying in g.

Proof. The multiplication map $N \times A \times K \to \mathrm{SL}_2(\mathbb{R})$ is continuous and open, so we need to show it is bijective. It is injective, because checking directly we see that

$$NA \cap K = \{1\} = N \cap A.$$

This map is surjective as follows. Let $g \in \mathrm{SL}_2(\mathbb{R})$, and let $z = g(i)$. Let $n_g = \begin{pmatrix} 1 & -\mathrm{Re}\, z \\ 0 & 1 \end{pmatrix} \in N$, so that $(n_g g)(i) = yi$. Let $a_g = \begin{pmatrix} 1/\sqrt{y} & 0 \\ 0 & \sqrt{y} \end{pmatrix} \in A$; then $(a_g n_g g)(i) = i$, so $a_g n_g g \in \mathrm{Stab}_{\mathrm{SL}_2(\mathbb{R})}(i) = K$, and peeling back we get $g \in NAK$, proving surjectivity. □

Remark 33.4.3. We have $AN = NA$, and we showed in the proof of Proposition 33.4.2 that NA acts transitively on $\mathbf{H}^2$ (by $z \mapsto a^2 z + b$). In section 34.6, we reinterpret this as providing a direct link between $\mathbf{H}^2$ and $\mathrm{SL}_2(\mathbb{R})$.

Lemma 33.4.4. *The group* $\mathrm{SL}_2(\mathbb{R})$ *is generated by the subgroups A, N, and the element* $\begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}$, *which acts on* $\mathbf{H}^2$ *by* $z \mapsto -1/z$.

Proof. Let $\begin{pmatrix} a & b \\ c & d \end{pmatrix} \in \mathrm{SL}_2(\mathbb{R})$. The lemma follows by performing row reduction on the matrix using the given generators. We find that if $c \neq 0$, then

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix} = \begin{pmatrix} 1 & a/c \\ 0 & 1 \end{pmatrix} \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} c & 0 \\ 0 & 1/c \end{pmatrix} \begin{pmatrix} 1 & d/c \\ 0 & 1 \end{pmatrix}$$

and if $c = 0$ then

$$\begin{pmatrix} a & b \\ 0 & 1/a \end{pmatrix} = \begin{pmatrix} a & 0 \\ 0 & 1/a \end{pmatrix} \begin{pmatrix} 1 & b/a \\ 0 & 1 \end{pmatrix}.$$

□

The subgroups N, A, K can be characterized by their traces; with a view to working on $\mathrm{PSL}_2(\mathbb{R})$, we consider the absolute traces:

$$|\mathrm{Tr}(K)| = [0, 2], \;\; |\mathrm{Tr}(A)| = [2, \infty), \;\text{ and } |\mathrm{Tr}(N)| = \{2\}.$$

Definition 33.4.5. An element $g \in \mathrm{PSL}_2(\mathbb{R})$ with $g \neq \pm 1$ is called **elliptic**, **hyperbolic**, or **parabolic** according to whether $|\mathrm{Tr}(g)| < 2$, $|\mathrm{Tr}(g)| > 2$, or $|\mathrm{Tr}(g)| = 2$.

Every nonidentity element $g \in \mathrm{PSL}_2(\mathbb{R})$ belongs to one of these three types (even though g need not belong to one of the subgroups N, A, K individually).

Lemma 33.4.6. *An element $g \in \mathrm{PSL}_2(\mathbb{R})$ is*

$$\begin{Bmatrix} \textit{elliptic} \\ \textit{hyperbolic} \\ \textit{parabolic} \end{Bmatrix} \textit{ if and only if } g \textit{ has } \begin{cases} \textit{a unique fixed point in } \mathbf{H}^2, \\ \textit{two fixed points on } \mathrm{bd}\,\mathbf{H}^2, \\ \textit{a unique fixed point on } \mathrm{bd}\,\mathbf{H}^2. \end{cases}$$

Proof. Let $g = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$ have $\det(g) = ad - bc = 1$. We look to solve the equation

$$\frac{az + b}{cz + d} = z$$

or equivalently $cz^2 + (d-a)z - b = 0$. The discriminant is $(d-a)^2 + 4bc = (a+d)^2 - 4 = \mathrm{Tr}(g)^2 - 4$. Therefore g is elliptic if and only if this discriminant is negative if and only if there is a unique root in $\mathbf{H}^2$; g is parabolic if and only if this discriminant is zero if and only if there is a unique root in $\mathrm{bd}\,\mathbf{H}^2$; and g is hyperbolic if and only if this discriminant is positive if and only if there are two roots in $\mathrm{bd}\,\mathbf{H}^2$. □

33.4.7. Let $g \in \mathrm{PSL}_2(\mathbb{R})$. If g is elliptic, then g acts by hyperbolic rotation in a neighborhood around its fixed point, as in Figure 33.4.8; every such element is conjugate to an element of K, fixing i. (Indeed, in the unit disc model with its fixed point as the center, an elliptic element acts literally by rotation in the disc; see section 33.7.)

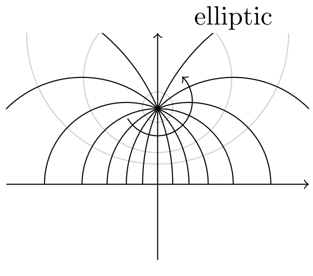


Figure 33.4.8: Action on $\mathbf{H}^2$ by an elliptic element

A hyperbolic element can be thought of as a translation along the geodesic between the two fixed points on $\mathrm{bd}\,\mathbf{H}^2$, as in Figure 33.4.9.

Moving these fixed points to $0, \infty$, every such element is conjugate to an element of A, acting by $z \mapsto a^2 z$ with $a \neq 1$, as in Figure 33.4.10.

Finally, a parabolic element should be thought of as a limit of the other two types, where correspondingly the fixed point tends to the boundary or the two fixed points move together; every such element is conjugate to an element of N, acting by translation $z \mapsto z + n$ for some $n \in \mathbb{R}$.

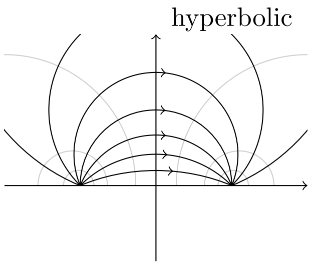


Figure 33.4.9: Action on H^2 by a hyperbolic element

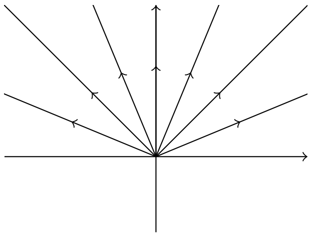


Figure 33.4.10: Action on H^2 by a hyperbolic element with fixed points 0, 1

Lemma 33.4.11. *For all $z, z' \in \mathsf{H}^2$, there exists $g \in \mathrm{PSL}_2(\mathbb{R})$ such that $gz, gz' \in \mathbb{R}_{>0}i$ are pure imaginary.*

Proof. The proof follows directly by using translations and scaling (33.4.1), requested in Exercise 33.5. □

33.5 Geodesics

In this section, we prove two important theorems: we describe geodesics, giving a formula for the distance, and we classify isometries.

Theorem 33.5.1. *The hyperbolic plane H^2 is a uniquely geodesic space. The unique geodesic passing through two distinct points $z, z' \in \mathsf{H}^2$ is a semicircle orthogonal to $\mathbb{R}$ or a vertical line, and*

$$\rho(z, z') = \log \frac{|z - \overline{z'}| + |z - z'|}{|z - \overline{z'}| - |z - z'|} \tag{33.5.2}$$

$$\cosh \rho(z, z') = 1 + \frac{|z - z'|^2}{2\,\mathrm{Im}(z)\,\mathrm{Im}(z')}. \tag{33.5.3}$$

Proof. We first prove the imaginary axis is a geodesic with $z, z' \in \mathbb{R}_{>0}i$. Let $\upsilon(t) = x(t) + iy(t) : z \to z'$ be a path; then

$$\begin{aligned}\ell(\upsilon) = \int_0^1 \frac{\sqrt{(\mathrm{d}x/\mathrm{d}t)^2 + (\mathrm{d}y/\mathrm{d}t)^2}}{y(t)}\,\mathrm{d}t \geq \int_0^1 \frac{\mathrm{d}y/\mathrm{d}t}{y(t)}\,\mathrm{d}t \\ = \log y(1) - \log y(0) = \log\left|\frac{z'}{z}\right|\end{aligned} \tag{33.5.4}$$

with equality if and only if $x(t) = 0$ identically and $\mathrm{d}y/\mathrm{d}t \geq 0$ for all $t \in [0,1]$. This is achieved for the path

$$\upsilon(t) = (|z|(1-t) + |z'|t)i;$$

so $\rho(z, z') = \log|z'/z|$, and the imaginary axis is the unique geodesic.

For arbitrary points $z, z' \in \mathbf{H}^2$, we apply Lemma 33.4.11. The statement on geodesics follows from the fact that the image of $\mathbb{R}_{>0}i$ under an element of $\mathrm{PSL}_2(\mathbb{R})$ is either a semicircle orthogonal to $\mathbb{R}$ or a vertical line (Exercise 33.6).

The formula (33.5.2) for the case $z, z' \in \mathbb{R}_{>0}i$ follows from (33.5.4) and plugging in along the imaginary axis; the general case then follows from the invariance of both $\rho(z, z')$ and

$$\log \frac{|z - \overline{z'}| + |z - z'|}{|z - \overline{z'}| - |z - z'|}$$

under $g \in \mathrm{PSL}_2(\mathbb{R})$, checked on the generators in Lemma 33.4.4 (Exercise 33.10). Finally, the formula (33.5.3) follows directly from formulas for hyperbolic cosine, requested in Exercise 33.11. □

Theorem 33.5.5. *We have*

$$\mathrm{Isom}(\mathbf{H}^2) \simeq \mathrm{PGL}_2(\mathbb{R})$$

and

$$\mathrm{Isom}^+(\mathbf{H}^2) \simeq \mathrm{PGL}_2^+(\mathbb{R}) \simeq \mathrm{PSL}_2(\mathbb{R}).$$

Proof. Let $Z = \{ti : t > 0\}$ be the positive part of the imaginary axis. By Theorem 33.5.1, Z is the unique geodesic through any two points of Z.

Let $\phi \in \mathrm{Isom}(\mathbf{H}^2)$. Then $\phi(Z)$ is a geodesic (33.2.5), so by Exercise 33.7, there exists $g \in \mathrm{PSL}_2(\mathbb{R})$ such that $g\phi$ fixes Z pointwise. Replacing ϕ by $g\phi$, we may suppose ϕ fixes Z pointwise.

Let $z = x + iy \in \mathbf{H}^2$ and $z' = x' + iy' = \phi(z)$. For all $t > 0$,

$$\rho(z, it) = \rho(\phi z, \phi(it)) = \rho(z', it).$$

Plugging this into the formula (33.5.3) for the distance, we obtain

$$(x^2 + (y-t)^2)y' = (x'^2 + (y'-t)^2)y.$$

Dividing both sides by t^2 and taking the limit as $t \to \infty$, we find that $y = y'$, and consequently that $x^2 = x'^2$ and $x = \pm x'$. The choice of sign $\pm$ varies continuously over the connected set $\mathbf{H}^2$ and so must be constant. Therefore $\phi(z) = z$ or $\phi(z) = -\overline{z}$ for all $z \in \mathbf{H}^2$. The latter generates $\mathrm{PGL}_2(\mathbb{R})$ over $\mathrm{PSL}_2(\mathbb{R})$ (33.3.10), and both statements in the theorem follow. □

33.6 Hyperbolic area and the Gauss–Bonnet formula

In this section, we consider hyperbolic area. We measure hyperbolic area by considering a small Euclidean rectangle whose sides are parallel to the axes at the point (x, y); the hyperbolic length of the sides are $\mathrm{d}x/y$ and $\mathrm{d}y/y$, and we obtain the hyperbolic area form from the product.

Definition 33.6.1. We define the **hyperbolic area form** by

$$\mathrm{d}A = \frac{\mathrm{d}x\,\mathrm{d}y}{y^2}.$$

For a subset $S \subseteq \mathbf{H}^2$, we define the **hyperbolic area** of S by

$$\mu(S) = \iint_S \mathrm{d}A$$

when this integral is defined.

Proposition 33.6.2. *The hyperbolic area is invariant under* $\mathrm{Isom}(\mathbf{H}^2)$.

Proof. We verify that the hyperbolic area form is invariant. We first check this for the orientation-reversing isometry

$$g(z) = g(x + iy) = -x + iy = -\overline{z};$$

visibly $\mathrm{d}(gA) = \mathrm{d}A$ in this case.

It suffices then to consider $g \in \mathrm{PSL}_2(\mathbb{R})$. Let $z = x + iy$ and let

$$w = g(z) = \frac{az + b}{cz + d} = u + iv,$$

with $ad - bc = 1$. By (33.3.9), $v = \dfrac{y}{|cz + d|^2}$. We compute that

$$\frac{\mathrm{d}g}{\mathrm{d}z}(z) = \frac{1}{(cz + d)^2}. \tag{33.6.3}$$

Now g is holomorphic; so by the Cauchy–Riemann equations, its Jacobian is given by

$$\left|\frac{\partial(u, v)}{\partial(x, y)}\right| = \left|\frac{\mathrm{d}g}{\mathrm{d}z}(z)\right|^2 = \frac{1}{|cz + d|^4};$$

therefore

$$\mathrm{d}(gA) = \frac{\mathrm{d}u\,\mathrm{d}v}{v^2} = \frac{\partial(u, v)}{\partial(x, y)}\frac{\mathrm{d}x\,\mathrm{d}y}{v^2} = \frac{1}{|cz + d|^4}\frac{|cz + d|^4}{y^2}\mathrm{d}x\,\mathrm{d}y = \mathrm{d}A. \qquad \square$$

A major role will be played in what follows by hyperbolic polygons, defined formally as follows.

33.6.4. Let $z, z' \in \mathbf{H}^{2*}$ be distinct. Then there is a unique geodesic in $\mathbf{H}^2$ whose closure in $\mathbf{H}^{2*}$ passes through z, z', and the segment from z to z' is denoted $[z, z']$.

Definition 33.6.5. A **hyperbolic polygon** is a connected, closed subset of $\mathbf{H}^{2*}$ whose boundary consists of finitely many geodesic **sides** of the form $[z, z']$ with z, z' **vertices**. A **hyperbolic triangle** is a hyperbolic polygon with three sides.

Definition 33.6.6. A subset $A \subseteq \mathbf{H}^2$ is **convex** if the geodesic segment between any two points in A lies inside A.

An example of a (convex) hyperbolic polygon is given in Figure 33.6.7.

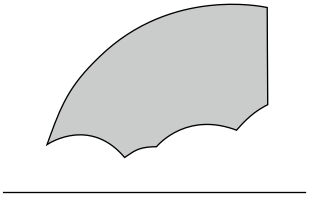


Figure 33.6.7: A hyperbolic polygon

A hyperbolic triangle is visibly convex; for more on convexity, see Exercises 33.8–33.9. The following key formula expresses the hyperbolic area of a triangle in terms of its angles.

Theorem 33.6.8 (Gauss–Bonnet formula). *Let T be a hyperbolic triangle with angles α, β, γ. Then*

$$\mu(T) = \pi - (\alpha + \beta + \gamma).$$

Proof. See e.g. Katok [Kat92, Theorem 1.4.2]. □

33.6.9. Let P be a convex hyperbolic polygon with n sides. By convexity, each side meets at each vertex a unique side, so P has n vertices with angles $\theta_1, \dots, \theta_n$. Triangulating P and applying the Gauss–Bonnet formula, we conclude that

$$\mu(P) = (n-2)\pi - (\theta_1 + \dots + \theta_n).$$

Remark 33.6.10. Theorem 33.6.8 is called the Gauss–Bonnet formula because it is closely related to the more general formula relating curvature to Euler characteristic. The simplest formula of this kind is

$$\int_X K \,\mathrm{d}A = 2\pi\chi(X) \tag{33.6.11}$$

for a compact Riemann surface X. The expression (33.6.11) is quite remarkable: it says that the total curvature of X is determined by its topology; if you flatten out a surface in one place, the curvature is forced to rise somewhere else. If instead one has a surface X with geodesic boundary, then the formula (33.6.11) becomes

$$\int_X K \,\mathrm{d}A + \sum_i (\pi - \theta_i) = 2\pi\chi(X)$$

where θ_i are the angles at the vertices. For a triangle X with constant curvature -1 and angles α, β, γ, we have $\int_X K\,\mathrm{d}A = -\mu(X)$ and $\chi(X) = V - E + F = 1$ (as for any polygon), so we find

$$-\mu(X) + 3\pi - (\alpha + \beta + \gamma) = 2\pi$$

and we recover Theorem 33.6.8.

33.7 Unit disc and Lorentz models

In this section, we consider two other models for the hyperbolic plane.

First, we consider the unit disc model.

Definition 33.7.1. The **hyperbolic unit disc** is the (open) unit disc

$$\mathbf{D}^2 = \{w \in \mathbb{C} : |w| < 1\}$$

equipped with the **hyperbolic metric**

$$\mathrm{d}s = \frac{2|\mathrm{d}w|}{1 - |w|^2}.$$

The hyperbolic unit disc $\mathbf{D}^2$ is also called the **Poincaré model** of planar hyperbolic geometry. The **circle at infinity** is the boundary

$$\operatorname{bd}\mathbf{D}^2 = \{w \in \mathbb{C} : |w| = 1\}.$$

33.7.2. For all $z_0 \in \mathbf{H}^2$, the maps

$$\begin{aligned} \phi_{z_0} : \mathbf{H}^2 &\xrightarrow{\sim} \mathbf{D}^2 & \phi_{z_0}^{-1} : \mathbf{D}^2 &\xrightarrow{\sim} \mathbf{H}^2 \\ z &\mapsto w = \frac{z - z_0}{z - \overline{z_0}} & w &\mapsto z = \frac{\overline{z_0}w - z_0}{w - 1} \end{aligned} \tag{33.7.3}$$

define a conformal equivalence between $\mathbf{H}^2$ and $\mathbf{D}^2$ with $z_0 \mapsto \phi(z_0) = 0$. A particularly nice choice is $z_0 = i$, giving

$$\phi(z) = \frac{z - i}{z + i}, \quad \phi^{-1}(w) = -i\frac{w + 1}{w - 1}. \tag{33.7.4}$$

The hyperbolic metric on $\mathbf{D}^2$ is the pushforward of (induced from) the hyperbolic metric on $\mathbf{H}^2$ via the identification (33.7.4) (Exercise 33.12). Ordinarily, one would decorate the pushforward metric, but because we will frequently move between the upper half-plane and unit disc as each has its advantage, we find it notationally simpler to avoid this extra decoration. The distance on $\mathbf{D}^2$ is

$$\begin{aligned} \rho(w, w') &= \log \frac{|1 - w\overline{w'}| + |w - w'|}{|1 - w\overline{w'}| - |w - w'|} \\ \cosh \rho(w, w') &= 1 + 2\frac{|w - w'|^2}{(1 - |w|^2)(1 - |w'|^2)} \end{aligned} \tag{33.7.5}$$

so that

$$\rho(w, 0) = \log \frac{1 + |w|}{1 - |w|} = 2\tanh^{-1}|w|. \tag{33.7.6}$$

The map ϕ (33.7.3) maps the geodesics in $\mathbf{H}^2$ to geodesics in $\mathbf{D}^2$, and as a Möbius transformation, maps circles and lines to circles and lines, preserves angles, and maps the real axis to the unit circle; therefore the geodesics in $\mathbf{D}^2$ are diameters through the origin and semicircles orthogonal to the unit circle, as in Figure 33.7.7.

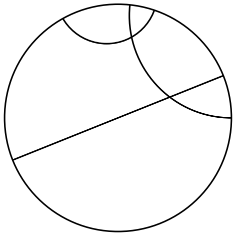


Figure 33.7.7: Hyperbolic geodesics

Accordingly, triangles from the upper half-plane map to triangles in the unit disc.

33.7.8. Via the map ϕ, the group $\mathrm{PSL}_2(\mathbb{R})$ acts on $\mathbf{D}^2$ as

$$\phi\,\mathrm{PSL}_2(\mathbb{R})\phi^{-1} = \mathrm{PSU}(1,1) = \left\{ \begin{pmatrix} a & b \\ \overline{b} & \overline{a} \end{pmatrix} \in \mathrm{M}_2(\mathbb{C}) : |a|^2 - |b|^2 = 1 \right\} /\{\pm 1\}.$$

Explicitly, an isometry of $\mathbf{D}^2$ is a map of the form

$$w \mapsto e^{i\theta}\left(\frac{w-a}{1-\overline{a}w}\right)$$

for $a \in \mathbb{C}$ with $|a| < 1$ and $\theta \in \mathbb{R}$. (A direct substitution can be used to give an alternate verification that these transformations are isometries of $\mathbf{D}^2$ with the hyperbolic metric.)

The orientation reversing isometry $g(z) = -\overline{z}$ on $\mathbf{H}^2$ acts by $g(w) = \overline{w}$ on $\mathbf{D}^2$ with the choice $p = i$ (Exercise 33.13).

The induced area on $\mathbf{D}^2$ is given by

$$\mathrm{d}A = \frac{4\,\mathrm{d}x\,\mathrm{d}y}{(1-x^2-y^2)^2}$$

for $w = x + yi$.

Second, we present the Lorentz model.

Definition 33.7.9. The **Lorentz metric** on $\mathbb{R}^3$ is the indefinite metric

$$\mathrm{d}s^2 = -\mathrm{d}t^2 + \mathrm{d}x^2 + \mathrm{d}y^2.$$

33.7.10. The indefinite Lorentz metric is associated to the quadratic form

$$q(t,x,y) = -t^2 + x^2 + y^2$$

in the natural way. Lengths in this metric can be positive or nonpositive. However, on the **hyperboloid**

$$t^2 - x^2 - y^2 = 1,$$

the metric becomes positive definite: a nonzero tangent vector to the hyperboloid has positive length (Exercise 33.16). The hyperboloid can be thought of as the sphere of radius i about the origin with respect to q; taking an imaginary radius shows that hyperbolic geometry is dual in some sense to spherical geometry, where $\mathbf{S}^2 \subseteq \mathbb{R}^3$ has real radius 1.

Definition 33.7.11. The **Lorentz hyperboloid** is the set

$$\mathbf{L}^2 = \{(t, x, y) \in \mathbb{R}^3 : q(t, x, y) = -1, t > 0\}$$

equipped with the Lorentz metric.

The Lorentz hyperboloid is the upper sheet of the (two-sheeted) hyperboloid; it is also called the **hyperboloid model** or the **Lorentz model** of planar hyperbolic geometry, and it can be visualized as in Figure 33.7.13. (The choice of signs has to do with the physics of spacetime.)

The map

$$\begin{aligned} \mathbf{L}^2 &\to \mathbf{D}^2 \\ (t, x, y) &\mapsto (x + iy)/(t + 1) \end{aligned} \tag{33.7.12}$$

is bijective and identifies the metrics on $\mathbf{L}^2$ and $\mathbf{D}^2$ (Exercise 33.15). Moreover, the map (33.7.12) maps geodesics in $\mathbf{D}^2$ to intersections of the hyperboloid with planes through the origin.

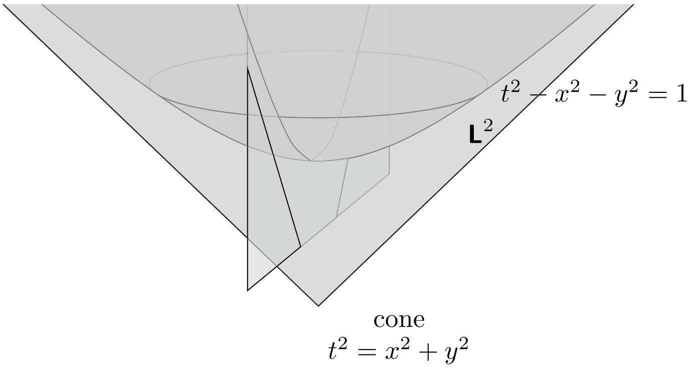


Figure 33.7.13: The hyperboloid model $\mathbf{L}^2$

By pullback, $\mathrm{Isom}^+(\mathbf{L}^2) \simeq \mathrm{PSL}_2(\mathbb{R})$. However, other isometries are also apparent: a linear change of variables that preserves the quadratic form q also preserves the Lorentz metric. Let

$$\mathrm{O}(2,1) = \{g \in \mathrm{GL}_3(\mathbb{R}) : g^t m g = m\}, \quad \text{where } m = \begin{pmatrix} -1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}.$$

Then

$$\mathrm{d}s^2 = v^t m v, \quad \text{where } v = (\mathrm{d}t, \mathrm{d}x, \mathrm{d}y)^t,$$

so if $g \in \mathrm{O}(2,1)$ then immediately $\mathrm{d}gs^2 = \mathrm{d}s^2$.

Next, if $g \in \mathrm{O}(2,1)$, then $g^t mg = m$ implies $\det(g) = \pm 1$. The elements of $\mathrm{O}(2,1)$ that map the hyperboloid to itself comprise the subgroup

$$\mathrm{SO}(2,1) = \{g \in \mathrm{O}(2,1) : \det(g) = 1\};$$

let $\mathrm{SO}^+(2,1) \leq \mathrm{SO}(2,1)$ be the further subgroup that maps the upper sheet of the hyperboloid to itself, the connected component of the identity.

Remark 33.7.14. We have proven that there is an isomorphism of Lie groups

$$\mathrm{SO}^+(2,1) \simeq \mathrm{PSL}_2(\mathbb{R}); \tag{33.7.15}$$

it corresponds to the isomorphism of Lie algebras $\mathfrak{so}_{2,1} \simeq \mathfrak{sl}_2$.

33.8 Riemannian geometry

The hyperbolic metric (33.3.4) is induced from a Riemannian metric as follows.

33.8.1. A **Riemannian metric** $\mathrm{d}s^2$ on an open set $U \subseteq \mathbb{R}^n$ is a function that assigns to each point $x \in U$ a (symmetric, positive definite) inner product on the tangent space $\mathrm{T}_x(U)$ at $x \in U$, varying differentiably. Such an inner product defines the length of a tangent vector $\| \ \|$, the angle between two tangent vectors, and the length element $\mathrm{d}s = \sqrt{\mathrm{d}s^2}$. In coordinates, we write

$$\mathrm{d}s^2 = \sum_{i,j} \eta_{ij}\,\mathrm{d}x_i\,\mathrm{d}x_j$$

for standard coordinates x_i on $\mathbb{R}^n$, and the matrix (η_{ij}) is symmetric, positive definite, and differentiable. The metric determines a volume formula as

$$\mathrm{d}V = \sqrt{\det \eta}\,\mathrm{d}x_1 \cdots \mathrm{d}x_n.$$

A Riemannian metric gives $U \subseteq \mathbb{R}^n$ the structure of a path metric space, as explained in 33.2.6: if $\upsilon\colon [0,1] \to U$ is continuously differentiable, then we define its length to be

$$\ell(\upsilon) = \int_\upsilon \mathrm{d}s = \int_0^1 \|\upsilon'(t)\|\,dt.$$

If $\phi\colon \mathbb{R}^k \to \mathbb{R}^n$ is continuously differentiable, the pullback metric $\phi^*(\mathrm{d}s^2)$ is defined by

$$\phi^*(\mathrm{d}s^2)(v,w) = \mathrm{d}s^2(\mathrm{D}f(v), \mathrm{D}f(w))$$

where $v, w \in \mathrm{T}_z(U)$ and D is the derivative map.

The language of 33.8.1 gives another way to interpret the hyperbolic metric on $\mathbf{H}^2$. This point of view extends to provide a description of the full isometry group $\mathrm{PSL}_2(\mathbb{R})$ as the unit tangent bundle of $\mathbf{H}^2$, as follows.

33.8.2. The tangent space to $\mathbf{H}^2$ at a point $z \in \mathbf{H}^2$ is $\mathrm{T}_z\mathbf{H}^2 \simeq \mathbb{C}$ and the tangent bundle

$$\mathrm{T}(\mathbf{H}^2) := \{(z,v) : z \in \mathbf{H}^2, v \in \mathrm{T}_z\mathbf{H}^2\}$$

is trivial (**parallelizable**), with $\mathrm{T}(\mathbf{H}^2) \simeq \mathbf{H}^2 \times \mathbb{C}$. The Riemannian metric on $\mathbf{H}^2$ is then defined by the metric on $\mathrm{T}_z\mathbf{H}^2$ given by

$$\langle v, w \rangle = \frac{v \cdot w}{(\operatorname{Im} z)^2}$$

for $v, w \in \mathrm{T}_z(\mathbf{H}^2)$, a rescaling of the usual inner product on $\mathbb{C}$ over $\mathbb{R}$. In particular, $\|v\| = |v|/(\operatorname{Im} z)$ for $v \in \mathrm{T}_z(\mathbf{H}^2)$. The angle between two geodesics at an intersection point $z \in \mathrm{T}_z\mathbf{H}^2$ is then defined to be the angle between their tangent vectors in $\mathrm{T}_z\mathbf{H}^2$; this notion of an angle coincides with the Euclidean angle measure.

The action of $\mathrm{PSL}_2(\mathbb{R})$ on $\mathbf{H}^2$ extends to an action on $\mathrm{T}(\mathbf{H}^2)$ in the expected way:

$$g(z, v) = \left(gz, \frac{\mathrm{d}g(z)}{\mathrm{d}z} v\right) = \left(\frac{az+b}{cz+d}, \frac{1}{(cz+d)^2} v\right).$$

Since isometries of $\mathbf{H}^2$ are differentiable, they act on the tangent bundle by differentials preserving the norm and angle, and therefore $\operatorname{Isom}(\mathbf{H}^2)$ acts conformally or anti-conformally on $\mathbf{H}^2$.

If we restrict to the **unit tangent bundle**

$$\mathrm{UT}(\mathbf{H}^2) := \{(z, v) \in \mathrm{T}(\mathbf{H}^2) : \|v\|_z^2 = 1\}$$

then we obtain a bijection

$$\begin{aligned} \mathrm{PSL}_2(\mathbb{R}) &\xrightarrow{\sim} \mathrm{UT}(\mathbf{H}^2) \\ g &\mapsto \left(gi, \frac{\mathrm{d}g}{\mathrm{d}z}(i) i\right) \end{aligned}$$

(Exercise 33.17).

Remark 33.8.3. The natural generalization of Euclid's geometry is performed on a Riemannian manifold X that is **homogeneous**, i.e., the isometry group $\operatorname{Isom}(X)$ acts transitively on X, as well as **isotropic**, i.e., $\operatorname{Isom}(X)$ acts transitively on frames (a basis of orthonormal tangent vectors) at a point. In this way, *homogeneous* says that every point "looks the same", and *isotropic* says that the geometry "looks the same in every direction" at a point. Taken together, these natural conditions are quite strong, and there are only three essentially distinct simply connected homogeneous and isotropic geometries in any dimension, corresponding to constant sectional curvatures zero, positive, or negative: these are Euclidean, spherical, and hyperbolic geometry, respectively. Put this way, the hyperbolic plane is the unique complete, simply connected Riemann surface with constant sectional curvature -1. For more on geometries in this sense, we encourage the reader to consult Thurston [Thu97].

To conclude this section, we briefly review a few facts from the theory of Riemannian manifolds.

33.8.4. A **(topological)** n**-manifold** is a (second-countable) Hausdorff topological space X locally homeomorphic to $\mathbb{R}^n$, i.e., for every $x \in X$, there exists an open neighborhood $U \ni x$ and a continuous map $\phi\colon U \hookrightarrow \mathbb{R}^n$ that is a homeomorphism

onto an open subset; the map $\phi\colon U \to \phi(U) \subseteq \mathbb{R}^n$ is called a **chart** (at $x \in X$), and an open cover of charts is called an **atlas**.

Two charts $\phi_1 : U_1 \to \mathbb{R}^n$ and $\phi_2 : U_2 \to \mathbb{R}^n$ are **(C^∞-)smoothly compatible** if the **transition map**

$$\phi_{12} = \phi_2\phi_1^{-1} : \phi_1(U_1 \cap U_2) \to \phi_2(U_1 \cap U_2)$$

(see Figure 33.8.5) is C^∞-smooth (i.e., has continuous partial derivatives of all orders). An atlas on a manifold is **smooth** on a manifold if all charts are smoothly compatible. A **smooth manifold** is a manifold equipped with a smooth atlas.

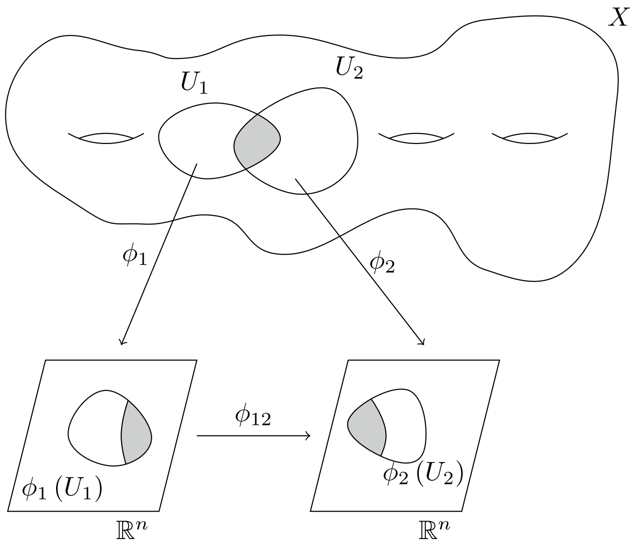


Figure 33.8.5: A manifold, by its atlas

A **morphism** of smooth manifolds is a continuous map $f\colon Y \to X$ such that for the atlases $\{(\phi_i, U_i)\}_i$ of X and $\{(\psi_j, V_j)\}_j$ of Y, each map

$$\phi_i f \psi_j^{-1} : \psi_j(V_j \cap f^{-1}(U_i)) \to \phi_i(f(V_j) \cap U_i)$$

is smooth. An **isomorphism** (**diffeomorphism**) of smooth manifolds is a bijective morphism $f\colon Y \xrightarrow{\sim} X$ such that f and f^{-1} are (C^∞-)smooth.

By the same definition as 33.8.1, we define a **Riemannian metric** on a smooth n-manifold.

One could similarly define C^k-smooth manifolds for any $1 \le k \le \infty$.

33.8.6. We similarly define a **complex** n**-manifold**, and morphisms between them, by replacing $\mathbb{R}$ by $\mathbb{C}$ and smooth by holomorphic in the definition of a smooth manifold. A **Riemann surface** is a complex 1-manifold. For further reference, see e.g. Donaldson [Don2011] or Miranda [Mir95].

A complex 1-manifold (Riemann surface) defines a smooth, orientable Riemannian 2-manifold by choosing the standard Euclidean metric on the complex plane;

conversely, a conformal structure on a smooth, oriented Riemannian 2-manifold determines a complex 1-manifold. In other words, the category of Riemann surfaces is equivalent to the category of smooth, orientable Riemannian 2-manifolds with conformal transition maps and with conformal morphisms.

Example 33.8.7. The field $\mathbb{C}$ of complex numbers is the "original" Riemann surface, and every open subset of $\mathbb{C}$ is a Riemann surface.

The simplest nonplanar example of a Riemann surface is the **Riemann sphere** $\mathbb{P}^1(\mathbb{C}) = \mathbb{C} \cup \{\infty\}$. The atlas on $\mathbb{P}^1(\mathbb{C})$ is given by the open sets

$$U_1 = \mathbb{P}^1(\mathbb{C}) \smallsetminus \{\infty\} = \mathbb{C} \text{ and } U_2 = \mathbb{P}^1(\mathbb{C}) \smallsetminus \{0\}$$

and charts $\phi_1 : U_1 \to \mathbb{C}$ by $\phi_1(z) = z$ and $\phi_2 : U_2 \to \mathbb{C}$ by $\phi_2(z) = 1/z$; the map $\phi_2\phi_1^{-1}(z) = 1/z$ is analytic on $\phi_1(U_1 \cap U_2) = \mathbb{C} \smallsetminus \{0\}$. Topologically, the Riemann sphere is the one-point compactification of $\mathbb{C}$, and becomes a sphere by stereographic projection.

Example 33.8.8. The inverse function theorem implies that if X is a smooth projective algebraic variety over $\mathbb{C}$, then $X(\mathbb{C})$ has the canonical structure of a compact, complex manifold.

Exercises

1. Show that the hyperbolic metric has the same topology as the Euclidean metric in two ways.

   (a) Show directly that open balls nest: for all $z \in \mathbf{H}^2$ and all $\epsilon > 0$, there exist $\eta_1, \eta_2 > 0$ such that
   $$\rho(z, w) < \eta_1 \Rightarrow |z - w| < \epsilon \Rightarrow \rho(z, w) < \eta_2$$
   for all $w \in \mathbf{H}^2$.

   (b) Show that the collection of Euclidean balls coincides with the collection of hyperbolic balls. *[Hint: applying an isometry, reduce to the case of balls around i and check this directly; it is perhaps even clearer moving to the unit disc model.]*

2. Check that in $\mathbb{R}^n$, the metric specified in (33.2.7)
   $$\ell(\upsilon) = \int_\upsilon \sqrt{x_1'(t)^2 + \cdots + x_n'(t)^2}\,\mathrm{d}t$$
   has lines as geodesics.

3. From differential geometry, the curvature of a Riemann surface with metric
   $$ds = \sqrt{f(x, y)\,\mathrm{d}x^2 + g(x, y)\,\mathrm{d}y^2}$$
   is given by the formula
   $$-\frac{1}{\sqrt{fg}}\left(\frac{\partial}{\partial x}\left(\frac{1}{\sqrt{f}}\frac{\partial\sqrt{g}}{\partial x}\right) + \frac{\partial}{\partial y}\left(\frac{1}{\sqrt{g}}\frac{\partial\sqrt{f}}{\partial y}\right)\right)$$

for suitably nice functions f, g. Using this formula, verify that the curvature of $\mathbf{H}^2$ and $\mathbf{D}^2$ is -1.

4. Consider $\mathbb{C}$ with the standard metric. Let

$$\mathrm{Isom}^h(\mathbb{C}) = \{g \in \mathrm{Isom}(\mathbb{C}) : g \text{ is holomorphic}\} \leq \mathrm{Isom}(\mathbb{C}).$$

Exhibit an isomorphism of groups

$$\mathrm{Isom}^h(\mathbb{C}) \simeq \left\{ \begin{pmatrix} a & b \\ 0 & 1 \end{pmatrix} \in \mathrm{GL}_2(\mathbb{C}) : |a| = 1 \right\}$$

and an isometry of $\mathbb{C}$ that is not holomorphic. *[Hint: An invertible holomorphic map* $\mathbb{C} \to \mathbb{C}$ *is of the form* $z \mapsto az + b$.]

▹ 5. Show that for every $z, z' \in \mathbf{H}^2$, there exists $g \in \mathrm{PSL}_2(\mathbb{R})$ such that $\mathrm{Re}\, gz = \mathrm{Re}\, gz' = 0$. *[Hint: Work in* $\mathbf{D}^2$.]

▹ 6. Show that the image of $\mathbb{R}_{>0}i$ under an element of $\mathrm{PSL}_2(\mathbb{R})$ is either a semicircle orthogonal to $\mathbb{R}$ or a vertical line. *[Hint: Look at the endpoints.]*

▹ 7. We consider the action of $\mathrm{PSL}_2(\mathbb{R})$ on geodesics in $\mathbf{H}^2$.

(a) Show that $\mathrm{PSL}_2(\mathbb{R})$ acts transitively on the set of geodesics in $\mathbf{H}^2$.

(b) Show that given $z, z' \in \mathbf{H}^2$, there exists $g \in \mathrm{PSL}_2(\mathbb{R})$ such that $gz = z'$ and such that g that maps the geodesic through z and z' to itself.

(c) Show that every isometry of $\mathbf{H}^2$ that maps a geodesic to itself and fixes two points on this geodesic is the identity.

▹ 8. Let $z_1, z_2 \in \mathbf{H}^2$ be distinct. Let

$$H(z_1, z_2) = \{z \in \mathbf{H}^2 : \rho(z, z_1) \leq \rho(z, z_2)\}$$

be the locus of points as close to z_1 as to z_2, and let $L(z_1, z_2) = \mathrm{bd}\, H(z_1, z_2)$. Show that $H(z_1, z_2)$ is a convex (Definition 33.6.6) half-plane, and that

$$L(z_1, z_2) = \{z \in \mathbf{H}^2 : \rho(z, z_1) = \rho(z, z_2)\}$$

is geodesic and equal to the perpendicular bisector of the geodesic segment from z_1 to z_2.

9. Show that a hyperbolic polygon is convex if and only if it is the intersection of finitely many half-planes $H(z_1, z_2)$ as in Exercise 33.8.

▹ 10. Show that the expression

$$\frac{|z - \overline{z'}| + |z - z'|}{|z - \overline{z'}| - |z - z'|}$$

with $z, z' \in \mathbf{H}^2$ is invariant under $g \in \mathrm{PSL}_2(\mathbb{R})$. *[Hint: check this on a convenient set of generators.]*

▹ 11. Show that

$$\cosh\log\frac{|z-\overline{z'}|+|z-z'|}{|z-\overline{z'}|-|z-z'|} = 1+\frac{|z-z'|^2}{2\,\mathrm{Im}(z)\,\mathrm{Im}(z')}$$

for all $z, z' \in \mathbf{H}^2$.

▹ 12. Verify that the hyperbolic metric on $\mathbf{D}^2$ is induced from the hyperbolic metric on $\mathbf{H}^2$ from the identification (33.7.4), as follows.

(a) Show that

$$\frac{2|\phi'(z)|}{1-|\phi(z)|^2} = \frac{1}{\mathrm{Im}\, z}.$$

(b) Let $w = \phi(z)$, and using part (a) conclude that

$$\frac{2\,|\mathrm{d}w|}{1-|w|^2} = \frac{|\mathrm{d}z|}{\mathrm{Im}\, z}.$$

13. Show that the orientation-reversing isometry $g(z) = -\overline{z}$ induces the map

$$(\phi g \phi^{-1})(w) = \overline{w}$$

on $\mathbf{D}^2$ via the conformal transformation $\phi\colon \mathbf{H}^2 \to \mathbf{D}^2$ in (33.7.4).

14. Show that the Iwasawa decomposition (Proposition 33.4.2) can be given explicitly as

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix} = \begin{pmatrix} 1 & (ac+bd)/r^2 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} 1/r & 0 \\ 0 & r \end{pmatrix} \begin{pmatrix} s & -t \\ t & s \end{pmatrix} \in NAK = \mathrm{SL}_2(\mathbb{R})$$

where $r = \sqrt{c^2+d^2}$, $s = d/r$, $t = c/r$.

15. Show that the map

$$\begin{aligned} \mathbf{L}^2 &\to \mathbf{D}^2 \\ (t, x, y) &\mapsto \frac{x+iy}{t+1} \end{aligned}$$

identifies the metrics on $\mathbf{L}^2$ and $\mathbf{D}^2$, via pullback.

▹ 16. Show that the Lorentz metric restricted to the hyperboloid is an honest (Riemannian) metric. *[Hint: Show that a tangent vector v at a point p satisfies $b(p, v) = 0$, where b is the bilinear form associated to q; then show that the orthogonal complement to p has signature $+2$.]*

▹ 17. Show that there is a bijection $\mathrm{PSL}_2(\mathbb{R}) \xrightarrow{\sim} \mathrm{UT}(\mathbf{H}^2)$ defined by the action of g on a fixed base point in $\mathrm{UT}(\mathbf{H}^2)$. *[Hint: Observe that elliptic elements rotate the tangent vector.]*

This chapter is licensed under the terms of the Creative Commons Attribution-NonCommercial 4.0 International License (http://creativecommons.org/licenses/by-nc/4.0/), which permits any noncommercial use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license and indicate if changes were made.

The images or other third party material in this chapter are included in the chapter's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the chapter's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.

Chapter 34

Discrete group actions

Our ongoing goal in this part of the text is to understand quotient spaces that locally look like (products of) hyperbolic spaces. In order to get off the ground, here we put the previous two chapters in a more general context, seeking to understand nice group actions on topological spaces and indicating how these fit in with more general notions in topology. Pathologies exist! Our goal in this chapter is to provide basic context (for further references see 34.5.4) before turning to the central case of interest: a discrete group acting properly on a locally compact, Hausdorff topological space.

34.1 Topological group actions

Group actions will figure prominently in what follows, so we set a bit of notation. There are many references for topological groups, including Arhangel'skii–Tkachenko [AT2008, Chapter 1] and McCarty [McC2011, Chapter V].

Let G be a group and let X be a set. Recall that a **left action** of G on X is a map

$$\begin{aligned} G \times X &\to X \\ (g, x) &\mapsto gx \end{aligned} \tag{34.1.1}$$

satisfying $1x = x$ and $(gg')x = g(g'x)$ for all $x \in X$ and all $g, g' \in G$. A right action is instead a map $X \times G \to X$ by $(x, g) \mapsto xg$ satisfying $x1 = x$ and $x(g'g) = (xg')g$ for all $x \in X$ and $g, g' \in g$. We will need to consider actions both on the left and the right; if not specified, a left action is assumed.

We will also sometimes write $G \circlearrowright X$ for an action of G on X.

Example 34.1.2. A group G acts on itself by left multiplication, the **(left) regular group action** of G. If $H \leq G$ is a subgroup, then H also acts on G by left multiplication. For example, if V is an $\mathbb{R}$-vector space with $\dim_{\mathbb{R}} V = n$, and $\Lambda \subseteq V$ is a (full) $\mathbb{Z}$-lattice in V, then $\Lambda \simeq \mathbb{Z}^n$ is a group and Λ acts on V by translation.

Another important and related example is the left action of G on the set of right cosets $X = G/H$ again by multiplication, namely

$$g(xH) = gxH \quad \text{for } g \in G \text{ and } xH \in G/H.$$

© The Author(s) 2021

J. Voight, *Quaternion Algebras*, Graduate Texts in Mathematics 288,
https://doi.org/10.1007/978-3-030-56694-4_34

Let G act on X. The G-**orbit** of $x \in X$ is $Gx = \{gx : g \in G\}$. The set of G-orbits forms the **quotient set** $G\backslash X = \{Gx : x \in X\}$, with a natural surjective **quotient map** $\pi \colon X \to G\backslash X$.

Remark 34.1.3. We write $G\backslash X$ for the quotient, as G acts on the left; for a right action, we write X/G, etc.

Example 34.1.4. A group G acts transitively on a nonempty set X if and only if $G\backslash X$ is a single point; in this case, we call X **homogeneous** under G. In particular, if $H \leq G$, then the action of G on G/H is transitive.

If $H \leq G$, then the quotient set $H\backslash G$ is the set of left cosets of H in G. For example, if $\Lambda = \mathbb{Z}^n \leq \mathbb{R}^n = V$, then $\Lambda\backslash V \simeq (\mathbb{R}/\mathbb{Z})^n$.

For $x \in X$, we define the **stabilizer** of x by $\mathrm{Stab}_G(x) = \{g \in G : gx = x\}$.

Definition 34.1.5. The action of G on X is **free** (and we say G acts **freely** on X) if $\mathrm{Stab}_G(x) = \{1\}$ for all $x \in X$, i.e., $gx = x$ implies $g = 1$ for all $x \in X$.

Definition 34.1.6. Let X', X be sets with an action of G. A map $f \colon X' \to X$ is G-**equivariant** if $f(gx') = g(f(x'))$ for all $x' \in X'$ and $g \in G$, i.e., the following diagram commutes:

$$\begin{array}{ccc} X' & \xrightarrow{f} & X \\ \downarrow{\scriptstyle g} & & \downarrow{\scriptstyle g} \\ X' & \xrightarrow{f} & X \end{array}$$

If $f \colon X' \to X$ is G-equivariant, then f induces a map

$$\begin{aligned} G\backslash X' &\to G\backslash X \\ Gx' &\mapsto Gf(x'), \end{aligned}$$

well-defined by the G-equivariance of f, and the following diagram commutes:

$$\begin{array}{ccc} X' & \xrightarrow{f} & X \\ \downarrow{\scriptstyle \pi'} & & \downarrow{\scriptstyle \pi} \\ G\backslash X' & \longrightarrow & G\backslash X \end{array} \tag{34.1.7}$$

Now topology enters. Let G be a topological group (Definition 12.2.1), a group with a topology in which the multiplication and inversion maps are continuous. Let X be a topological space, and let G act on X. We want to consider only those actions in which the topology on G and on X are compatible.

Definition 34.1.8. The action of G on X is **continuous** if the map $G \times X \to X$ is continuous.

Example 34.1.9. The left regular action of a group on itself is continuous—indeed, combined with continuity of inversion (and existence of the identity), this is the very definition of a topological group.

Lemma 34.1.10. *Suppose G has the discrete topology. Then an action of G on X is continuous if and only if for all $g \in G$ the left-multiplication map*

$$\begin{aligned} \lambda_g : X &\to X \\ x &\mapsto gx \end{aligned}$$

is continuous; and when this holds, each λ_g is a homeomorphism.

Proof. Exercise 34.2. ◻

From now on, suppose G acts continuously on X; more generally, whenever G is a topological group acting on a topological space X, we will implicitly suppose that the action is continuous.

34.1.11. The quotient $G\backslash X$ is equipped with the **quotient topology**, so that the quotient map $\pi \colon X \to G\backslash X$ is continuous: a subset $V \subseteq G\backslash X$ is open if and only if $\pi^{-1}(V) \subseteq X$ is open.

The projection π is an open map, which is to say if $U \subseteq X$ is open then $\pi(U) = GU \subseteq G\backslash X$ is open: if U is open then $\pi^{-1}(\pi(U)) = \bigcup_{g \in G} gU$ is open, so $\pi(U)$ is open by definition of the topology.

34.1.12. If G acts continuously on X, then the topologies on G and X are related by this action. In particular, for all $x \in X$, the natural map

$$\begin{aligned} G &\to Gx \subseteq X \\ g &\mapsto gx \end{aligned}$$

is continuous (it is the restriction of the action map to $G \times \{x\}$). Let $K = \operatorname{Stab}_G(x)$. Then this map factors naturally as

$$G/K \xrightarrow{\sim} Gx \tag{34.1.13}$$

where we give G/K the quotient topology; then (34.1.13) then a bijective continuous map, a topological upgrade of the orbit–stabilizer theorem. The map (34.1.13) need not always be a homeomorphism (Exercise 34.5), but we will see below that it becomes a homeomorphism under further nice hypotheses (Exercise 34.6, Proposition 34.4.11).

In order to work concretely with the quotient $G\backslash X$, it is convenient to choose representatives of each orbit as follows. We write cl for topological closure and int for topological interior.

Definition 34.1.14. A **fundamental set** for $G \circlearrowright X$ is a subset $\text{⌑} \subseteq X$ such that:

(i) $\operatorname{cl}(\operatorname{int}(\text{⌑})) = \text{⌑}$;

(ii) $G\text{⌑} = X$; and

(iii) $\operatorname{int}(\text{⌑}) \cap \operatorname{int}(g\text{⌑}) = \emptyset$ for all $1 \neq g \in G$.

The condition (i) ensures our basic intuition about tilings (and avoids fundamental sets that contain an extraneous number of isolated points); condition (ii) says that ⌑ tiles X; and condition (iii) shows that the tiles only overlap along the boundary, and there is no redundancy in the interior. If there is a fundamental set for $G \circlearrowright X$, then the action is faithful.

34.1.15. Let ⌑ $\subseteq X$ be a fundamental set for $G \circlearrowright X$. Then G induces an equivalence relation on ⌑, and $G\backslash$⌑ $\xrightarrow{\sim} G\backslash X$ is a bijection.

Remark 34.1.16. In chapter 37, we place further restrictions on a fundamental set to ensure that they retain good properties, calling such a set a *fundamental domain* (Definition 37.1.11).

34.2 ▹ Summary of results

We pause to provide a quick summary of the results in this chapter for the special case of discrete subgroups of $\mathrm{PSL}_2(\mathbb{R})$ (proven in Theorem 34.5.1 and Proposition 34.7.2). The reader who is willing to accept the theorem below, and will stick to this case, can profitably skip the rest of this chapter.

The group $\mathrm{PSL}_2(\mathbb{R})$ has a natural topology from the metric on $\mathrm{SL}_2(\mathbb{R}) \subseteq \mathrm{M}_2(\mathbb{R})$ (see 34.6.1): intuitively, two matrices in $\mathrm{PSL}_2(\mathbb{R})$ are close if after a choice of sign all of their entries are close.

Theorem 34.2.1. *Let* $\Gamma \leq \mathrm{PSL}_2(\mathbb{R})$ *be a subgroup and equip* Γ *with the subspace topology. Then the following are equivalent:*

(i) Γ *is discrete;*
(ii) *For all* $z \in \mathbf{H}^2$*, we have* $\#\,\mathrm{Stab}_\Gamma(z) < \infty$ *and there exists an open neigborhood* $U \ni z$ *such that* $\gamma U \cap U \neq \emptyset$ *implies* $\gamma \in \mathrm{Stab}_\Gamma(z)$*;*
(iii) *For all compact subsets* $K \subseteq \mathbf{H}^2$*, we have* $K \cap \gamma K \neq \emptyset$ *for only finitely many* $\gamma \in \Gamma$*; and*
(iv) *For all* $z \in \mathbf{H}^2$*, the orbit* $\Gamma z \subseteq \mathbf{H}^2$ *is discrete and* $\#\,\mathrm{Stab}_\Gamma(z) < \infty$.

Moreover, if these equivalent conditions hold, then the quotient $\Gamma\backslash\mathbf{H}^2$ *is Hausdorff, and the quotient map* $\pi\colon \mathbf{H}^2 \to \Gamma\backslash\mathbf{H}^2$ *is a local isometry at all points* $z \in \mathbf{H}^2$ *with* $\mathrm{Stab}_\Gamma(z) = \{1\}$.

A discrete subgroup $\Gamma \leq \mathrm{PSL}_2(\mathbb{R})$ is called a **Fuchsian group**.

34.3 Covering space and wandering actions

Throughout the remainder of this chapter, let X be a Hausdorff topological space with a (continuous) action of a Hausdorff topological group G.

Definition 34.3.1. We say the action of G on X is a **covering space action** if for all $x \in X$, there exists an open neighborhood $U \ni x$ such that $gU \cap U = \emptyset$ for all $g \in G$ with $g \neq 1$.

34.3.2. If $G \circlearrowright X$ is a covering space action, then the quotient map $\pi\colon X \to G\backslash X$ is a **local homeomorphism**, i.e., for every $x \in X$, there exists an open neighborhood $U \ni x$ such that $\pi|_U : U \to \pi(U) \subseteq G\backslash X$ is a homeomorphism. A local homeomorphism need not conversely be a covering space map.

If G acts by a covering space action, then G acts freely on X. This is too strong a hypothesis on the group actions we will consider in the rest of this book, so we need to look for something weaker. So we consider the following.

Definition 34.3.3. We say that the action of G is **wandering** if for all $x \in X$, there exists an open neighborhood $U \ni x$ such that $gU \cap U \neq \emptyset$ for only finitely many $g \in G$.

Example 34.3.4. If G is a finite group, then every action of G is wandering.

34.3.5. If the action of G is wandering, then for all $x \in X$, the orbit $Gx \subseteq X$ is closed and discrete.

Wandering actions generalize covering space actions, and can be equivalently characterized, as follows.

Lemma 34.3.6. *The following are equivalent:*

(i) *The action of G is wandering; and*
(ii) *For all $x \in X$ we have $\#\operatorname{Stab}_G(x) < \infty$, and there exists an open neighborhood $U \ni x$ such that $gU \cap U \neq \emptyset$ implies $g \in \operatorname{Stab}_G(x)$.*

If G acts freely, then these are further equivalent to:

(iii) *The action of G is by a covering space action.*

Proof. The implication (ii) $\Rightarrow$ (i) is immediate; we prove the converse. Let U be a neighborhood of $x \in X$ such that $gU \cap U \neq \emptyset$ for only finitely many $g \in G$. We have $\#\operatorname{Stab}_G(x) < \infty$ since $g \in \operatorname{Stab}_G(x)$ implies $x \in gU \cap U$. Let

$$\{g \in G : gU \cap U \neq \emptyset \text{ and } gx \neq x\} = \{g_1, \dots, g_n\}. \tag{34.3.7}$$

Since X is Hausdorff, for all i there exist open neighborhoods $V_i, W_i \subseteq X$ of $x, g_i x$, respectively, such that $V_i \cap W_i = \emptyset$. Replacing V_i with $V_i \cap U$ (still with $x \in V_i$ and $V_i \cap W_i = \emptyset$), we may suppose that $V_i \subseteq U$. Since G acts continuously, there exists an open neighborhood $W_i' \subseteq X$ of x such that $g_i W_i' \subseteq W_i$. Let $U_i := V_i \cap W_i'$. Since $V_i \subseteq U$ we have $U_i \subseteq U$ for all i. Further, $x \in U_i$ and

$$U_i \cap g_i U_i \subseteq U_i \cap g_i W_i' \subseteq V_i \cap W_i = \emptyset. \tag{34.3.8}$$

We claim that $U' = \bigcap_i U_i$ has the desired property in (ii). Suppose that $gU' \cap U' \neq \emptyset$ for some $g \in G$. Since $U_i \subseteq U$ for all i we have $U' \subseteq U$, so $gU \cap U \neq \emptyset$; then by (34.3.7), either $gx = x$ or $g = g_i$ for some i. We cannot have $g = g_i$, since this would imply $g_i U' \cap U' \subseteq g_i U_i \cap U_i = \emptyset$, contradicting (34.3.8). So $gx = x$, and $g \in \operatorname{Stab}_G(x)$ as claimed.

Finally, if G acts freely, then $\operatorname{Stab}_G(x)$ is trivial for all x, whence the equivalence (ii) $\Leftrightarrow$ (iii). ∎

34.3.9. Suppose the action of G is wandering. We recall Lemma 34.3.6(ii). At a point x with open neighborhood $U \ni x$ and finite stabilizer $\mathrm{Stab}_G(x)$, we can replace U by $\bigcap_{g \in \mathrm{Stab}_G(x)} gU$ so that $U \ni x$ is an open neighborhood on which $\mathrm{Stab}_G(x)$ acts (see Exercise 34.12). Then the projection map factors as

$$\pi|_U : U \to \mathrm{Stab}_G(x)\backslash U \to G\backslash X$$

and the latter map $\mathrm{Stab}_G(x)\backslash U \to G\backslash X$ is a homeomorphism onto its image; we say π is a **local homeomorphism modulo stabilizers**. If the action of G is free, then we recover 34.3.2.

Remark 34.3.10. If G has the discrete topology and the condition in Lemma 34.3.6(ii) holds, then some authors call the action of G **properly discontinuous**. This is probably because G is then as broken ("discontinuous") as possible: G has the discrete topology, and we should be able to find neighborhoods that pull apart the action of G. (Klein [Kle79, p. 321] uses the term *discontinuous* because "points that are 'equivalent' with respect to [the group] are separated".) This nomenclature is strange because we still want the action to be continuous, just by a discrete group. Adding to the potential confusion is the issue that different authors give different definitions of "properly discontinuous" depending on their purposes; most of these can be seen to be equivalent under the right hypotheses on the space, but not all. We avoid this term.

It turns out that a wandering action is too weak a property in this level of generality for us to work with. However, it is close, and we will shortly see that it suffices with additional hypotheses on the space X.

Remark 34.3.11. Let X be a topological space, and let G be a set of continuous maps $X \to X$. Then there is a natural map $G \hookrightarrow X^X$ defined by $g \mapsto (gx)_x$. We give X^X the compact-open topology and G the subspace topology, so a subbasis of the topology on G is given by

$$V(K, U) = \{f \in G : f(K) \subseteq U\}$$

for $K \subseteq X$ compact and $U \subseteq X$ open.

If X is Hausdorff and locally compact, then the compact-open topology on G is the weakest topology (smallest, fewest open sets) for which the map $G \times X \to X$ is continuous (also called an **admissible** topology on G) [McC2011, §VII, pp. 171–172]. Under the hypotheses of Exercise 34.4, this implies that the topology of pointwise convergence and the compact-open topology coincide.

34.4 Hausdorff quotients and proper group actions

In this section we define proper group actions; to motivate this definition, we first ask for conditions that imply that a quotient space is Hausdorff. Throughout this section and the next, let X be a Hausdorff topological space and let G be a Hausdorff topological group acting continuously on X.

Lemma 34.4.1. *The following are equivalent:*

(i) *The quotient $G\backslash X$ is Hausdorff;*

(ii) *If* $Gx \neq Gy \in G\backslash X$*, then there exist open neighborhoods* $U \ni x$ *and* $V \ni y$ *such that* $gU \cap V = \emptyset$ *for all* $g \in G$*; and*

(iii) *The image of the action map*

$$\begin{aligned} G \times X &\to X \times X \\ (g, x) &\mapsto (x, gx) \end{aligned} \tag{34.4.2}$$

is closed.

Proof. The implication (i) ⇔ (ii) follows directly from properties of the quotient map: the preimage of open neighborhoods separating Gx and Gy under the continuous projection map have the desired properties, and conversely the pushforward of the given neighborhoods under the open projection map separate Gx and Gy.

To conclude, we prove (i) ⇔ (iii). We use the criterion that a topological space is Hausdorff if and only if the diagonal map has closed image. The continuous surjective map $\pi \colon X \to G\backslash X$ is open, so the same is true for

$$\pi \times \pi \colon X \times X \to (G\backslash X) \times (G\backslash X).$$

Therefore the diagonal $G\backslash X \hookrightarrow (G\backslash X) \times (G\backslash X)$ is closed if and only if its preimage is closed in $X \times X$. But this preimage consists exactly of the orbit relation

$$\{(x, x') \in X \times X : x' = gx \text{ for some } g \in G\},$$

and this is precisely the image of the action map (34.4.2). □

The conditions Lemma 34.4.1(i)–(ii) can sometimes be hard to verify, so it is convenient to have a condition that implies Lemma 34.4.1(iii); this definition will seek to generalize the situation when G is compact. First, we make a definition.

Definition 34.4.3. Let $f \colon X \to Y$ be a continuous map.

(a) We say $f \colon X \to Y$ is **quasi-proper** if $f^{-1}(K)$ is compact for all compact $K \subseteq Y$.
(b) We say f is **proper** if f is quasi-proper and closed (the image of every closed subset is closed).

Example 34.4.4. If X is compact, then every continuous map $f \colon X \to Y$ is proper because f is closed and if $K \subseteq Y$ is compact, then K is closed, so $f^{-1}(K) \subset X$ is closed hence compact, since X is compact.

Lemma 34.4.5. *Suppose that* Y *is locally compact and Hausdorff, and let* $f \colon X \to Y$ *be continuous and quasi-proper. Then* X *is locally compact, and* f *is proper.*

Proof. For the first statement, cover Y with open relatively compact sets $U_i \subseteq K_i$; then $V_i = f^{-1}(U_i)$ is an open cover of X by relatively compact sets.

Next, we claim that f is in fact already proper; that is to say, we show that f is closed. Let $W \subseteq X$ be a closed set and consider a sequence $\{y_n\}_n$ from $f(W)$ with $y_n \to y$. Let K be a compact neighborhood of y containing $\{y_n\}$; taking a subsequence, we may suppose all $y_n \in K$. Let $x_n \in f^{-1}(y_n) \cap W$ be primages. Since f is quasi-

proper, we have $f^{-1}(K)$ is compact. Suppose for a moment that $f^{-1}(K)$ is sequentially compact (for example, if X is second countable or metrizable). Then again taking a subsequence, we may suppose that $x_n \to x$ with $x \in W$ since W is closed. By continuity, $f(x_n) \to f(x) = y$, so $f(W)$ is closed. To avoid the extra hypothesis that $f^{-1}(K)$ is sequentially compact, replace the sequence $\{y_n\}$ with a net; the argument proceeds identically. □

Remark 34.4.6. There is an alternate characterization of proper maps as follows: a continuous map $f : X \to Y$ is proper if and only if the map $f \times \mathrm{id} : X \times Z \to Y \times Z$ is closed for every topological space Z. See 34.5.4 for more discussion.

Partly motivated by Lemma 34.4.1(iv), we make the following definition.

Definition 34.4.7. The action of G on X is **proper** (G acts **properly** on X) if the **action map**

$$\begin{aligned} \lambda : G \times X &\to X \times X \\ (g, x) &\mapsto (x, gx) \end{aligned} \tag{34.4.8}$$

is proper.

Proposition 34.4.9. *If G is compact, then every (continuous) action of G on (a Hausdorff space) X is proper.*

Proof. Let $K \subseteq X \times X$ be compact; then K is closed (because X is Hausdorff). Let K_1 be the projection of K onto the first factor. Then K_1 is compact, and $\lambda^{-1}(K)$ is a closed subset of the compact set $G \times K_1$, so it is compact. This shows that the action map is quasi-proper. Finally, the action map is closed. We factor the map as

$$\begin{aligned} G \times X &\to G \times X \times X \to X \times X \\ (g, x) &\mapsto \quad (g, x, gx) \mapsto (x, gx); \end{aligned}$$

the first map is the graph of a continuous map to a Hausdorff space and is closed (Exercise 34.9); the second (projection) map is closed, as G is compact (by a standard application of the tube lemma). Therefore the composition of these maps is closed. □

Example 34.4.10. If G is a finite discrete group, then G acts properly by Proposition 34.4.9.

Proper actions have many of the properties we need.

Proposition 34.4.11. *Let G act properly on X. Then the following are true.*

(a) *$G\backslash X$ is Hausdorff.*
(b) *The orbit $Gx \subseteq X$ is closed for all $x \in X$.*
(c) *The natural map*

$$\begin{aligned} \iota_x : G/\operatorname{Stab}_G(x) &\to Gx \\ g &\mapsto gx \in X \end{aligned}$$

is a homeomorphism.

(d) *The group* $\mathrm{Stab}_G(x)$ *is compact for all* $x \in X$.

Proof. For part (a), by Lemma 34.4.1, it is enough to note that by definition the image of the action map λ in (34.4.8) is closed. Part (b) follows in the same way, as

$$Gx \simeq \{x\} \times Gx = \lambda(G \times \{x\}).$$

This also implies part (c) (cf. 34.1.12): the map ι_x is bijective and continuous, and it is also closed (whence a homeomorphism) since ι_x is a factor of the closed map $G \to Gx$.

Finally, for part (d), let $\lambda : G \times X \to X \times X$ be the action map and let $x \in X$. Then by definition that $\lambda^{-1}(x, x) = \mathrm{Stab}_G(x) \times \{x\} \simeq \mathrm{Stab}_G(x)$, so by definition $\mathrm{Stab}_G(x)$ is compact. □

34.5 Proper actions on a locally compact space

When X is locally compact, our central case of interest, there are several equivalent characterizations of a proper discrete action $G \circlearrowright X$. For more about proper group actions and covering spaces, see Lee [Lee2011, Chapter 12].

Recall our running assumption that X and G are Hausdorff.

Theorem 34.5.1. *Suppose that* X *is locally compact and let* G *act (continuously) on* X. *Then the following are equivalent:*

(i) G *is discrete and acts properly on* X;
(ii) *For all compact subsets* $K \subseteq X$, *we have* $K \cap gK \neq \emptyset$ *for only finitely many* $g \in G$;
(iii) *For all compact subsets* $K, L \subseteq X$, *we have* $K \cap gL \neq \emptyset$ *for only finitely many* $g \in G$; *and*
(iv) *For all* $x, y \in X$, *there exist open neighborhoods* $U \ni x$ *and* $V \ni y$ *such that* $U \cap gV \neq \emptyset$ *for only finitely many* $g \in G$.

Moreover, if X *is a locally compact metric space and* G *acts by* isometries, *then these are further equivalent to:*

(v) *The action of* G *on* X *is wandering; and*
(vi) *For all* $x \in X$, *the orbit* $Gx \subseteq X$ *is discrete and* $\#\,\mathrm{Stab}_G(x) < \infty$.

Proof. First, we show (i) ⇒ (ii). Let $\lambda : G \times X \to X \times X$ be the action map. Let $K \subseteq X$ be compact. Then

$$\lambda^{-1}(K \times K) = \{(g, x) \in G \times X : x \in K, gx \in K\}$$

is compact by definition. The projection of $\lambda^{-1}(K \times K)$ onto G is compact, and since G is discrete, this projection is finite and includes all $g \in G$ such that $K \cap gK \neq \emptyset$.

Next we show (ii) ⇔ (iii): The implication (ii) ⇐ (iii) is immediate, and conversely we apply (ii) to the compact set $K \cup L$ to conclude

$$K \cap gL \subseteq (K \cup L) \cap g(K \cup L) \neq \emptyset$$

for only finitely many $g \in G$.

Next we show (ii) $\Rightarrow$ (iv). For all $x \in X$, since X is locally compact there is a compact neighborhood $K \supseteq U \ni x$, with U open and K compact. If $U \cap gU \neq \emptyset$ then $K \cap gK \neq \emptyset$ and this happens for only finitely many $g \in G$.

Finally, we show (iv) $\Rightarrow$ (i). We first show that the action map is quasi-proper, and conclude that it is proper by Lemma 34.4.5. Let $K \subseteq X \times X$ be compact. By (iv), for all $(x, y) \in K$, there exist open neighborhoods $U \ni x$ and $V \ni y$ such that the set

$$W = \{g \in G : gU \cap V \neq \emptyset\}$$

is finite. The set $U \times V \ni (x, y)$ is an open neighborhood of $(x, y) \in K$, and so the collection of these neighborhoods ranging over $(x, y) \in K$ is an open cover of K, so finitely many $U_i \times V_i \ni (x_i, y_i)$ suffice, and with corresponding sets $\#W_i < \infty$. Let $W = \bigcup_i W_i \subseteq G$. Let $K_1 \subseteq X$ be the projection of K onto the first coordinate. We claim that $\lambda^{-1}(K) \subseteq W \times K_1$. Indeed, if $\lambda(g, x) = (x, gx) \in K$ then $x \in K_1$ and $(x, gx) \in U_i \times V_i$ for some i, so $gx \in gU_i \cap V_i$ and $g \in W_i$, and thus $(g, x) \in W \times K_1$. Since $\#W < \infty$ and K_1 is compact, $W \times K_1$ is compact; and then since K is compact, K is closed so $\lambda^{-1}(K) \subseteq W \times K_1$ is also closed, hence compact.

To conclude that G is discrete, we argue as follows. For all $x \in X$, the orbit $Gx \subseteq X$ is discrete: taking $U = V$ and a neighborhood $U \ni x$ with $U \cap gU \neq \emptyset$ for only finitely many $g \in G$, we see that $U \cap Gx$ is finite so Gx is discrete (as X is Hausdorff). By Proposition 34.4.11(d), the map

$$G/\operatorname{Stab}_G(x) \to Gx$$

is a homeomorphism for all $x \in X$. Therefore, $\operatorname{Stab}_G(x)$ (the preimage of x) is an open, finite (Hausdorff) neighborhood of 1; but then $\operatorname{Stab}_G(x)$ is discrete, and transporting we conclude that the topological group G has an open cover by discrete sets, and thus G is discrete. This completes the equivalence (i)–(iv).

The implication (iv) $\Rightarrow$ (v) holds in all cases: taking $x = y$, the neighborhood $U \cap V$ is as required in the definition of a wandering action. The implication (v) $\Rightarrow$ (vi) also holds in all cases from 34.3.5 and Lemma 34.3.6.

To conclude, we show (vi) $\Rightarrow$ (ii) under the extra hypothesis that X is a metric space with G acting by isometries. Assume for purposes of contradiction that there exist infinitely many $g_n \in G$ such that $K \cap g_n K \neq \emptyset$, and accordingly let $x_n \in K$ with $g_n x_n \in K$. The points x_n accumulate in K, so we may suppose $x_n \to x \in K$; by taking a further subsequence, we may suppose also that $g_n x_n \to y \in K$. We then claim that the set $\{g_n x\}_n$ accumulates near y. Since $\#\operatorname{Stab}_G(x) < \infty$, we may suppose without loss of generality that the points $g_n x$ are all distinct. Then, given $\epsilon > 0$,

$$\rho(g_n x, y) \leq \rho(g_n x, g_n x_n) + \rho(g_n x_n, y) = \rho(x, x_n) + \rho(g_n x_n, y) < \frac{\epsilon}{2} + \frac{\epsilon}{2} = \epsilon,$$

for n sufficiently large, so $g_n x \to y$. Let $h_n = g_{n+1}^{-1} g_n \in G$. By the Cauchy criterion,

$$\rho(h_n x, x) = d(g_n x, g_{n+1} x) < \epsilon$$

for n sufficiently large. Since $h_n x \neq x$ for all n, this contradicts that the orbit Gx is discrete, having no limit points. $\square$

Remark 34.5.2. The hypothesis "X is a metric space with G acting by isometries" providing the equivalent condition Theorem 34.5.1(v) is necessary: see Exercise 34.11.

34.5.3. From Lemma 34.3.6 and the implication Theorem 34.5.1(v) $\Rightarrow$ (i), we see that proper actions generalize covering space actions when X is locally compact metric space and G acts by isometries. In fact, a more general statement is true: if G is a discrete group with a covering space action on X such that $G\backslash X$ is Hausdorff, then G acts properly on X. The (slightly involved) proof in general is requested in Exercise 34.16.

Remark 34.5.4. Bourbaki discusses proper maps [Bou60, Chapter I, §10] and more generally groups acting properly on topological spaces [Bou60, Chapter III, §§1,4]; the definition of proper is equivalent to ours as follows. Let $f\colon X \to Y$ be continuous, and say f is **Bourbaki proper** to mean that $f \times \mathrm{id}\colon X \times Z \to Y \times Z$ is closed for every topological space Z. If f is Bourbaki proper, then f is proper [Bou60, Chapter I, §10, Proposition 6]. In the other direction, if f is proper then f is closed and $f^{-1}(y)$ is compact for all $y \in Y$, and this implies that f is Bourbaki proper [Bou60, Chapter I, §10, Theorem 1].

34.6 Symmetric space model

In this section, before proceeding with our treatment of discrete group actions in our case of interest, we pause to give a very important way to think about hyperbolic space in terms of symmetric spaces. The magical formulas in hyperbolic geometry beg for a more conceptual explanation: what is their provenance? Although it is important for geometric intuition to begin with a concrete model of hyperbolic space and ask about its isometries directly, from this point of view it is more natural to instead *start* with the desired group and have it act on itself in a natural way.

34.6.1. Let $G = \mathrm{SL}_2(\mathbb{R})$. As a matrix group, G comes with a natural metric. The space $\mathrm{M}_2(\mathbb{R}) \simeq \mathbb{R}^4$ has the usual structure of a metric space, with

$$\|g\|^2 = a^2 + b^2 + c^2 + d^2, \quad \text{if } g = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in \mathrm{M}_2(\mathbb{R}).$$

We give $\mathrm{SL}_2(\mathbb{R}) \subset \mathrm{M}_2(\mathbb{R})$ the subspace metric and $\mathrm{PSL}_2(\mathbb{R})$ the quotient metric. Intuitively, in this metric $g, h \in \mathrm{PSL}_2(\mathbb{R})$ are close if there exist matrices representing g, h (corresponding to a choice of sign) with all four entries of the matrix close in $\mathbb{R}$.

34.6.2. Recall from 34.1.12 that if G acts (continuously and) transitively on X, then for all $x \in X$, the natural map $g \mapsto gx$ gives a continuous bijection

$$G/\operatorname{Stab}_G(x) \xrightarrow{\sim} Gx = X.$$

Let $X = \mathbf{H}^2$ be the hyperbolic plane and let $G = \mathrm{SL}_2(\mathbb{R})$. Then G acts transitively on X. The stabilizer of $x = i$ is the subgroup $K = \operatorname{Stab}_G(x) = \mathrm{SO}(2) \leq \mathrm{SL}_2(\mathbb{R})$, so there

is a continuous bijection

$$\begin{aligned} G/K = \mathrm{SL}_2(\mathbb{R})/\,\mathrm{SO}(2) &\xrightarrow{\sim} \mathbf{H}^2 = X \\ gK &\mapsto gi. \end{aligned} \tag{34.6.3}$$

From the Iwasawa decomposition (Proposition 33.4.2), it follows that

$$\mathrm{SL}_2(\mathbb{R})/\,\mathrm{SO}(2) \simeq NA. \tag{34.6.4}$$

In fact, the map (34.6.3) is a homeomorphism. To prove this, we observe the following beautiful equation: for $g \in \mathrm{SL}_2(\mathbb{R})$,

$$\|g\|^2 = 2\cosh \rho(i, gi). \tag{34.6.5}$$

This formula follows directly from the formula (33.5.3) for distance; the calculation is requested in Exercise 34.17. It follows that the map $G \to X$ is open, and thus (34.6.3) is a homeomorphism. In fact, by (34.6.5), if we reparametrize the metric on either $\mathrm{SL}_2(\mathbb{R})/\,\mathrm{SO}(2)$ or $\mathbf{H}^2$ by the appropriate factor involving the hyperbolic cosine, the map (34.6.3) becomes an isometry.

We conclude this section with a view to a more general setting where the above situation applies.

34.6.6. Let G be a connected, Hausdorff, locally compact topological group. We recall (section 29.3) that G has a *Haar measure*, a Borel measure μ that is left-translation invariant (so $\mu(gA) = \mu(A)$ for all Borel sets $A \subseteq G$ and $g \in G$). The Haar measure is unique up to scaling, with the Haar measure on $G = \mathbb{R}^n$ the usual Lebesgue measure.

G has a maximal compact subgroup $K \leq G$, unique up to conjugation in G, and the quotient $X = G/K$ is homeomorphic to Euclidean space—in particular, X is contractible.

A **lattice** $\Gamma \leq G$ is a discrete subgroup such that $\mu(\Gamma\backslash G) < \infty$. A lattice Γ acts properly on X by left multiplication.

Remark 34.6.7. More generally, a **(globally) symmetric space** is a space of the form G/K where G is a Lie group and $K \leq G$ a maximal compact subgroup. Alternatively, it can be defined as a space where every point has a neighborhood where there is an isometry of order 2 fixing the point. For more reading on the theory of symmetric spaces, and the connection to differential geometry and Lie groups, see the book by Helgason [Hel2001].

34.7 Fuchsian groups

We now specialize to our case of interest and consider the group $\mathrm{PSL}_2(\mathbb{R})$ acting by isometries on the geodesic space $\mathbf{H}^2$. A gentle introduction to the geometry of discrete groups is provided by Beardon [Bea95], with a particular emphasis on Fuchsian groups and their fundamental domains—in the notes at the end of each chapter are further bibliographic pointers. See also Jones–Singerman [JS87].

Lemma 34.7.1. *Let* $\Gamma \leq \mathrm{SL}_2(\mathbb{R})$. *Then the following are equivalent:*

(i) Γ *is discrete;*

(ii) *If* $\gamma_n \in \Gamma$ *and* $\gamma_n \to 1$, *then* $\gamma_n = 1$ *for almost all n; and*

(iii) *For all* $M \in \mathbb{R}_{>0}$, *the set* $\{\gamma \in \Gamma : \|\gamma\| \leq M\}$ *is finite.*

Proof. The equivalence (i) $\Leftrightarrow$ (ii) is requested in Exercise 34.15. The implication (i) $\Leftrightarrow$ (iii) follows from the fact that the ball of radius M in $\mathrm{SL}_2(\mathbb{R})$ is a compact subset of $\mathrm{M}_2(\mathbb{R})$, and a subset of a compact set is finite if and only if it is discrete. Slightly more elaborately, a sequence of matrices with bounded norm has a subsequence where the entries all converge; since the determinant is continuous, the limit exists in $\mathrm{SL}_2(\mathbb{R})$ so Γ is not discrete. □

In particular, we find from Lemma 34.7.1 that a discrete subgroup of $\mathrm{SL}_2(\mathbb{R})$ is countable.

Proposition 34.7.2. *Let* $\Gamma \leq \mathrm{PSL}_2(\mathbb{R})$ *be a subgroup (with the subspace topology). Then* Γ *has a wandering action on* $\mathbf{H}^2$ *if and only* Γ *is discrete.*

Proof. The implication $\Rightarrow$ is a consequence of Theorem 34.5.1(v) $\Rightarrow$ (i). Conversely, suppose that Γ is discrete; we show that Theorem 34.5.1(vi) holds: that for all $x \in X$, the orbit $Gx \subseteq X$ is discrete and $\#\,\mathrm{Stab}_G(x) < \infty$.

First we show that the stabilizer of a point is finite. We may work in the unit disc $\mathbf{D}^2$ and take the point to be $w = 0 \in \mathbf{D}^2$, as in 33.7.8. The stabilizer of $w = 0$ in $\mathrm{SU}(1,1)$ is $\mathrm{SO}(2) \simeq \mathbb{R}/(2\pi)\mathbb{Z}$, so its stabilizer in Γ is a discrete subgroup of the compact group $\mathrm{SO}(2)$ and is necessarily finite (indeed, cyclic).

Next we show that orbits of Γ on $\mathbf{H}^2$ are discrete. We apply the identity (34.6.5). This identity with Lemma 34.7.1 shows that the orbit Γi is discrete. But for all $z \in \mathbf{H}^2$, there exists $\phi \in \mathrm{PSL}_2(\mathbb{R})$ such that $\phi(i) = z$, and conjugation by ϕ induces an isomorphism $\Gamma \xrightarrow{\sim} \phi^{-1}\Gamma\phi$ of topological groups. Since

$$\rho(z, gz) = \rho(\phi(i), g\phi(i)) = \rho(i, (\phi^{-1}g\phi)i)$$

applying the above argument to $\phi^{-1}\Gamma\phi$ shows that the orbit Γz is discrete. This concludes the proof.

Alternatively, here is a self-contained proof that avoids the slightly more involved topological machinery. We again work in the unit disc $\mathbf{D}^2$. First we prove ($\Leftarrow$). Since Γ is discrete, there is an ϵ-neighborhood $U \ni 1$ with $U \subseteq \mathrm{PSU}(1,1)$ such that $U \cap \Gamma = \{1\}$; therefore, if

$$\gamma = \begin{pmatrix} a & b \\ \overline{b} & \overline{a} \end{pmatrix} \in \Gamma \setminus \{1\}$$

then $|b| > \epsilon$ or (without loss of generality) $|a-1| > \epsilon$. We claim that in either case

$$|\gamma(0)| = \left|\frac{b}{a}\right| > \epsilon,$$

and thus the orbit is discrete. Indeed, if $|b| > \epsilon$, then since $|a| < 1$ anyway immediately $|b/a| > \epsilon$; if $|a-1| > \epsilon$ then $|a| < 1-\epsilon$ so $|a|^2 < 1-\epsilon^2$ and $1/|a|^2 > 1+\epsilon^2$, giving

$$\left|\frac{b}{a}\right|^2 = \frac{1-|a|^2}{|a|^2} > (1+\epsilon^2) - 1 = \epsilon^2.$$

For ($\Rightarrow$), suppose that Γ is not discrete; then there is a sequence $\gamma_n \in \Gamma \setminus \{1\}$ of elements such that $\gamma_n \to 1$. Therefore, for all $z \in \mathbf{H}^2$, we have $\gamma_n z \to z$ and $\gamma_n z = z$ for only finitely many n, so every neighborhood of z contains infinitely many distinct points $\gamma_n z$. $\square$

With this characterization, we make the following important definition.

Definition 34.7.3. A **Fuchsian group** is a discrete subgroup of $\mathrm{PSL}_2(\mathbb{R})$.

A Fuchsian group Γ acts by orientation-preserving isometries on $\mathbf{H}^2$; this action is proper and wandering by Theorem 34.5.1.

34.7.4. A Fuchsian group $\Gamma \leq \mathrm{PSL}_2(\mathbb{R})$ is **elementary** if there is a nonempty Γ-invariant set in $\mathbf{H}^2 \cup \mathrm{bd}\,\mathbf{H}^2$ that contains at most two points. Equivalently, an elementary group is a cyclic subgroup or a (possibly) dihedral group—in particular, an elementary group is virtually abelian (has a finite index, abelian subgroup). The elementary groups are easy to analyze, but their inclusion into theorems about more general Fuchsian groups can cause problems; and so in general we are only interested in non-elementary groups.

Non-elementary Fuchsian groups Γ are categorized by the set of limit points $L(\Gamma) \subseteq \mathrm{bd}\,\mathbf{H}^2$ of Γz with $z \in \mathbf{H}^2$. If $L(\Gamma) = \mathrm{bd}\,\mathbf{H}^2$, then Γ is said to be a Fuchsian group **of the first kind**; otherwise Γ is **of the second kind**, and $L(\Gamma)$ is a nowhere-dense perfect subset of $\mathrm{bd}\,\mathbf{H}^2$, topologically a Cantor set. We will see later that if Γ has quotient with finite hyperbolic area, then Γ is finitely generated of the first kind.

34.8 Riemann uniformization and orbifolds

Our understanding of group actions has an important consequence for the classification of Riemann surfaces, and we pause (again) to provide this application.

First, we have the important structural result.

Theorem 34.8.1 (Riemann uniformization theorem). *Every (connected and) simply connected Riemann surface H is isomorphic to either the Riemann sphere $\mathbb{P}^1(\mathbb{C})$, the complex plane $\mathbb{C}$, or the hyperbolic plane $\mathbf{H}^2$.*

A consequence of Riemann uniformization is as follows.

34.8.2 The universal cover $\widetilde{X}$ of a compact Riemann surface X is simply connected, so by the theory of covering spaces, X is a quotient $X \simeq \Gamma\backslash\widetilde{X}$ where Γ is the **fundamental group** of X, a subgroup of isometries of $\widetilde{X}$ acting by a covering space action.

When $\widetilde{X} = \mathbb{P}^1(\mathbb{C})$, the only possible group Γ (acting freely) is trivial. When $\widetilde{X} = \mathbb{C}$, by classification one sees that the only Riemann surfaces of the form $X = \mathbb{C}/\Gamma$ are the

plane $X = \mathbb{C}$, the punctured plane $\mathbb{C}^\times \simeq \mathbb{C}/\langle u \rangle$ with $u \in \mathbb{C}^\times$, and complex tori $\mathbb{C}/\Lambda$ where $\Lambda \subset \mathbb{C}$ is a lattice with $\Lambda \simeq \mathbb{Z}^2$. We will embark on a classification of these tori up to isomorphism by their j-invariants in section 40.1.

All other Riemann surfaces are **hyperbolic** with $\widetilde{X} = \mathbf{H}^2$, and so are of the form $X = \Gamma\backslash\mathbf{H}^2$ with $\Gamma \leq \mathrm{PSL}_2(\mathbb{R})$ a torsion-free Fuchsian group.

Remark 34.8.3. Klein and Poincaré conjectured the uniformization theorem for algebraic curves over $\mathbb{C}$, with rigorous proofs were given by Poincaré: see Gray [Gray94].

Finally, before departing our topological treatment, we consider quotients of manifolds by the (continuous) action of a group. As we have seen, it is quite restrictive to suppose that this group action is free: we will still want to take quotients by such groups. Accordingly, we need to model not spaces that are locally modelled by $\mathbb{R}^n$ but those that are locally modelled by the quotient of $\mathbb{R}^n$ by a finite group.

Definition 34.8.4 An n**-orbifold** X is a (second-countable) Hausdorff topological space that is locally homeomorphic to a quotient $G\backslash\mathbb{R}^n$ with G a finite group acting (continuously). An **atlas** for an orbifold X is the data

(i) An open cover $\{U_i\}_{i\in I}$ of **charts** U_i closed under finite intersection; and
(ii) For each $i \in I$, an open subset $V_i \subseteq \mathbb{R}^n$ equipped with the (continuous) action of a finite group $G_i \circlearrowright V_i$, and a homeomorphism

$$\phi_i \colon U_i \xrightarrow{\sim} G_i\backslash V_i$$

satisfying the **atlas axiom**: for all $U_i \subseteq U_j$, there exists an injective group homomorphism $f_{ij} : G_i \hookrightarrow G_j$ and a G_i-equivariant map $\psi_{ij} : V_i \hookrightarrow V_j$ satisfying $\phi_j \circ \psi_{ij} = \phi_i$ (see Figure 34.8.5).

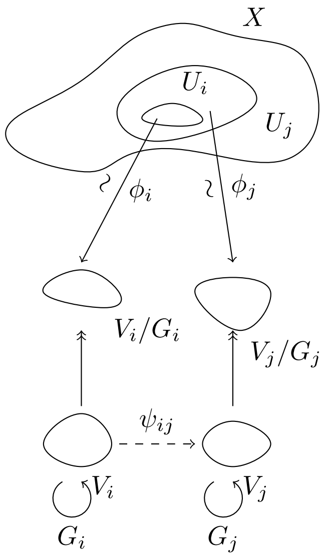


Figure 34.8.5: An orbifold, by its atlas

Orbifolds were introduced by Thurston [Thu97, Chapter 13], who adds a wealth of motivation and examples; see also the surveys by Scott [Sco83, §2] and Gordon [Gor2012] as well as the chapter by Ratcliffe [Rat2006, Chapter 13].

34.8.6 We can further ask that the transition maps f_{ij} in an atlas be smooth to get a **smooth** orbifold, preserve a G_i-Riemannian metric to get a **Riemann** orbifold, etc.; replacing $\mathbb{R}^n$ by $\mathbb{C}^n$ and smooth by holomorphic, we similarly define a **complex n-orbifold**, locally modelled on the quotient $G\backslash\mathbb{C}^n$ with G a finite group acting holomorphically.

Definition 34.8.7 Let X be an n-orbifold.

(a) A point $z \in X$ such that there exists a chart $U_i \ni z$ with group $G_i \neq \{1\}$ fixing z is called an **orbifold point** of X, with **stabilizer group** (or **isotropy group**) G_i; the set of orbifold points of X is called the **orbifold set** of X.
(b) If $z \in U_i$ is an isolated orbifold point and its stabilizer group is cyclic, we call z a **cone point**.

34.8.8 The prototypical example of an orbifold is the quotient of $\mathbb{C}$ by a finite group of rotations. Such a group is necessarily cyclic (as a finite subgroup of $\mathbb{C}^\times$) of some order $m \geq 2$; the quotient is a *cone*, a fundamental set for the action being a segment with angle $2\pi/m$, and the fixed point is a cone point of order m. The cone is homeomorphic to $\mathbb{R}^2$ but it is not isometric: away from the cone point, this space is locally isometric to $\mathbb{R}^2$, but at the cone point the angle is less than 2π, so shortest paths that do not start or end at the cone point never go through the cone point.

34.8.9 Let X be a manifold and let G be a finite group acting (continuously) on X such that action of G is wandering (Definition 34.3.3). We define an orbifold $[X/G]$ as follows: by Lemma 34.3.6 and 34.3.9, we can refine an atlas of X to one consisting of open neighborhoods U_i on which $G \circlearrowright U_i$ acts, and we make this into an orbifold atlas by taking $G_i = G$ for each i; the atlas axiom is tautologically satisfied.

When X is smooth, complex, Riemann, etc., we ask that G act diffeomorphically, holomorphically, etc., to obtain an orbifold with the same properties.

A full, suitable definition of the category of orbifolds—in particular, morphisms between them—is more subtle than it may seem. In this text, we will be primarily interested in an accessible and well-behaved class of orbifolds obtained as the quotient of a manifold.

Definition 34.8.10 An orbifold is **good** if is of the form $[X/G]$, i.e., it arises as the quotient of a manifold by a finite group.

34.8.11 The quotient $[X/G]$ of a Riemannian manifold X by a discrete group G of isometries acting properly is a good Riemann orbifold.

Example 34.8.12 A complex 1-orbifold is good if and only if it has a branched cover by a Riemann surface. By Exercise 34.18, the only complex 1-orbifolds that are not good are the **teardrop**, a sphere with one cone point, and the **football**, a sphere with two cone points of different orders.

34.8.13 Good (topological) compact, oriented 2-orbifolds admit a classification (extending the usual classification of surfaces by genus) up to homeomorphism by their **signature** $(g; e_1, \dots, e_k)$, where g is the genus of the underlying topological surface and the $e_1, \dots, e_k$ are the orders of the (necessarily cyclic) nontrivial stabilizer groups.

34.8.14 Putting these two pieces together, now let $\Gamma \leq \mathrm{PSL}_2(\mathbb{R})$ be a Fuchsian group. Then the discrete group $\Gamma \circlearrowright \mathbf{H}^2$ acts properly. Then $\Gamma\backslash\mathbf{H}^2$ has the structure of a good complex 1-orbifold, by the main theorem (Theorem 34.2.1(ii)).

Remark 34.8.15. From certain topological points of view, especially with an eye towards generalizations, an orbifold is best understood as a topological groupoid (a point of view first noticed by Haefliger [Hae84, §4]): the objects of the category are elements of the disjoint union of the charts U_i, and a morphism from $u_i \in U_i$ to $u_j \in U_j$ is the germ of a local homeomorphism that commutes with the projections. For more on the categorical perspective of orbifolds as groupoids, see Moerdijk [Moe2002] and Moerdijk–Pronk [MP1997].

Exercises

Unless otherwise specified, let $G \circlearrowright X$ be an action of a group G on a set X.

1. Show that the quotient map $\pi\colon X \to G\backslash X$ is defined by a universal property: if $f\colon X \to Z$ is a G-equivariant map where Z has a trivial G-action, then f factors uniquely through π, i.e., there exists a unique map $h\colon G\backslash X \to Z$ making the diagram

$$\begin{array}{ccc} X & \xrightarrow{\ f\ } & Z \\ {\scriptstyle \pi}\downarrow & \nearrow{\scriptstyle h} & \\ G\backslash X & & \end{array}$$

commute.

▹ 2. Prove Lemma 34.1.10, in the following form. Let G be a topological group acting on a topological space X.

 (a) Show that if the action is continuous, then for all $g \in G$ the map $X \to X$ by $x \mapsto gx$ is continuous (therefore, a homeomorphism).
 (b) Show the converse of (a) if G is discrete.

3. Let G be a topological group acting continuously on a topological space X. Show that the orbits of G are closed ($Gx \subseteq X$ is closed for all $x \in X$) if and only if $G\backslash X$ is T_1 (see Exercise 12.2).
4. Let X be a metric space. Then $\mathrm{Isom}(X)$ has naturally the topology of **pointwise convergence**, as follows. There is an embedding

$$\begin{aligned} \mathrm{Isom}(X) &\hookrightarrow X^X = \prod_{x \in X} X \\ g &\mapsto (g(x))_{x \in X}. \end{aligned}$$

The product X^X has the product topology, and so Isom(X) (and every space of maps from X to X) has an induced subspace topology. A basis of open sets for Isom(X) in this topology consists of finite intersections of open balls

$$V(g; x, \epsilon) = \{h \in \mathrm{Isom}(X) : \rho(g(x), h(x)) < \epsilon\}.$$

Equip the group $G = \mathrm{Isom}(X)$ with the topology of pointwise convergence.

(a) Show that G is a topological group.
(b) Show that G acts continuously on X.

5. Let $G = \mathbb{Z}$ be given the discrete topology, and let $G \circlearrowright X = \mathbb{R}/\mathbb{Z}$ act by $x \mapsto x + na \in \mathbb{R}/\mathbb{Z}$ for $n \in \mathbb{Z}$ for $a \in \mathbb{R} - \mathbb{Q}$. Show that this action is free and continuous, and show that for all $x \in X$ the map (34.1.13)

$$\begin{aligned} G/\operatorname{Stab}_G(x) = G &\to Gx \\ g &\mapsto gx \end{aligned}$$

is (continuous and bijective but) *not* a homeomorphism, giving $Gx \subseteq X$ the subspace topology.

▹ 6. Let G act (continuously and) transitively on X. Suppose that G, X are (Hausdorff and) locally compact, and suppose further that G has a countable base of open sets. Let $x \in X$ and let $K = \operatorname{Stab}_G(x)$. Show that the natural map $G/K \to X$ is a homeomorphism.

7. Let $G \circlearrowright X$ be a free and wandering action, and let U be an open set such that $gU \cap U = \emptyset$ for all $g \neq 1$. Show that the map $G \times U \to \pi^{-1}(\pi(U))$ is a homeomorphism and the restriction $\pi\colon G \times U \to \pi(U) \simeq U$ is a (split) covering map.

8. Let X be (Hausdorff and) locally compact, let $x \in X$, and let $U \ni x$ be an open neighborhood. Show that there exists an open neighborhood $V \ni x$ such that $K = \mathrm{cl}(V) \subseteq U$ is compact.

▹ 9. Let X, Y be (Hausdorff) topological spaces, let $f\colon X \to Y$ be a continuous map, and let

$$\begin{aligned} \mathrm{gr}(f) : X &\to X \times Y \\ x &\mapsto (x, f(x)) \end{aligned}$$

be the graph of f. Show that f is a closed map.

10. One way to weaken the running hypothesis that X is Hausdorff in this chapter is to instead assume only that X is **locally Hausdorff**: every $x \in X$ has an open neighborhood $U \ni x$ such that U is Hausdorff.
Show that a weakened version of Lemma 34.3.6(i) ⇒ (ii) is not true with only the hypothesis that X is locally Hausdorff: that is, exhibit a locally Hausdorff topological space X with a (continuous) wandering action of a group G such that $\pi\colon X \to G\backslash X$ is *not* a local homeomorphism, and so Lemma 34.3.6(ii) does not hold. *[Hint: Let X be the bug-eyed line and $G \simeq \mathbb{Z}/2\mathbb{Z}$ acting by $x \mapsto -x$* on $\mathbb{R}^\times$ *and swapping points in the doubled origin.]*

11. Let $G = \mathbb{Z}$ and let $G \hookrightarrow X = \mathbb{R}^2 \smallsetminus \{(0,0)\}$ act by $n \cdot (x, y) = (2^n x, y/2^n)$. In other words, G is the group of continuous maps $X \to X$ generated by $(x, y) \mapsto (2x, y/2)$.

    (a) Show that the action of G on X is free and wandering.
    (b) Show that the quotient $G\backslash X$ is not Hausdorff.
    (c) Let $K = \{(t, 1-t) : t \in [0, 1]\}$. Then K is compact. Show that $K \cap gK \neq \emptyset$ for infinitely many $g \in G$. [So Theorem 34.5.1(v) holds but (ii) does not, and in particular that the action of G is not proper. Can you see this directly from the definition of proper?]

12. Let X be a Hausdorff topological space with a continuous action of a Hausdorff topological group G. Suppose that the action of G is wandering. Show that for all $x \in X$, there is an open neighborhood $U \ni x$ such that the finite group $\mathrm{Stab}_G(x)$ acts on U (i.e., $gU \subseteq U$ for all $g \in \mathrm{Stab}_G(x)$).
13. Show that a subgroup $\Gamma \leq \mathbb{R}^n$ is discrete if and only if $\Gamma = \mathbb{Z}v_1 + \cdots + \mathbb{Z}v_m$ with $v_1, \dots, v_m \in \Gamma$ linearly independent over $\mathbb{R}$. As a consequence, show that $\Gamma \leq \mathbb{R}^n$ is a lattice if and only if Γ is discrete with $m = n$.
14. Exhibit an injective group homomorphism $\mathrm{SO}(n) \hookrightarrow \mathrm{SO}(n+1)$ and a homeomorphism
$$\mathsf{S}^n \simeq \mathrm{SO}(n+1)/\,\mathrm{SO}(n),$$
where $\mathsf{S}^n = \{x \in \mathbb{R}^{n+1} : \|x\|^2 = 1\}$ is the n-dimensional sphere, analogous to (34.6.3).

▹ 15. Let G be a topological group with a countable system of fundamental open neighborhoods of $1 \in G$ (for example, this holds if G is metrizable). Show that G is discrete if and only if whenever $\{g_n\}_n$ is a sequence from G with $g_n \to 1$, then $g_n = 1$ for all but finitely many n.

16. Let G be a discrete group with a (continuous) covering space action on a Hausdorff space X such that $G\backslash X$ is Hausdorff. Show that G acts quasi-properly on X.

▹ 17. Show that for $g \in \mathrm{SL}_2(\mathbb{R})$,
$$\|g\|^2 = 2\cosh\rho(i, gi)$$
(cf. 34.6.1). *[Hint: Use the formula (33.5.2).]*

18. (a) Show that a compact, complex 1-orbifold is good if and only if it has a branched cover by a compact Riemann surface.
    (b) Use the Riemann–Hurwitz theorem to show that the only compact, complex 1-orbifolds that are not good are the teardrop (a sphere with one cone point) and the football (a sphere with two cone points of different orders).
    (c) Show that every finitely generated discrete group of isometries of a simply connected Riemann surface with compact quotient has a torsion free subgroup of finite index. *[Hint: find a torsion free subgroup of finite index by avoiding the finitely many conjugacy classes of torsion in Γ.]* Use this to give another proof of (b).

19. Show that the stabilizer group of an orbifold point is well-defined up to group isomorphism, independent of the chart.
20. The group $SL_2(\mathbb{R})$ acts on $\mathbb{P}^1(\mathbb{R}) = \mathbb{R} \cup \{\infty\}$ by linear fractional transformations. Show that $G = SL_2(\mathbb{Z}) \leq SL_2(\mathbb{R})$ is discrete, but G does not act properly on $\mathbb{P}^1(\mathbb{R})$. [So discrete groups can act on locally compact spaces without necessarily acting properly.]

This chapter is licensed under the terms of the Creative Commons Attribution-NonCommercial 4.0 International License (http://creativecommons.org/licenses/by-nc/4.0/), which permits any noncommercial use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license and indicate if changes were made.

The images or other third party material in this chapter are included in the chapter's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the chapter's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.

Chapter 35

Classical modular group

In this chapter, we introduce the classical modular group $\mathrm{PSL}_2(\mathbb{Z}) \leq \mathrm{PSL}_2(\mathbb{R})$, a discrete group acting on the upper half-plane that has received extensive study because of the role it plays throughout mathematics. We examine the group in detail via a fundamental domain and conclude with some applications to number theory. This chapter will serve as motivation and example for the generalizations sought later in this part of the text.

There are very many references for the classical modular group, including Apostol [Apo90, Chapter 2], Diamond–Shurman [DS2005, Chapter 2], and Serre [Ser73, Chapter VII].

35.1 ▹ The fundamental set

Definition 35.1.1. The **classical modular group** is the subgroup of $\mathrm{PSL}_2(\mathbb{R})$ defined by

$$\mathrm{PSL}_2(\mathbb{Z}) = \left\{ \gamma = \begin{pmatrix} a & b \\ c & d \end{pmatrix} : a, b, c, d \in \mathbb{Z}, ad - bc = 1 \right\} / \{\pm 1\}.$$

The group $\mathrm{PSL}_2(\mathbb{Z})$ acts faithfully on the upper half-plane $\mathbf{H}^2$ by linear fractional transformations; equipping $\mathbf{H}^2$ with the hyperbolic metric, this action is by orientation-preserving isometries.

Since $\mathbb{Z} \subseteq \mathbb{R}$ is discrete, so too is $\mathrm{SL}_2(\mathbb{Z}) \subseteq \mathrm{M}_2(\mathbb{Z}) \subseteq \mathrm{M}_2(\mathbb{R})$ discrete and therefore $\mathrm{PSL}_2(\mathbb{Z}) \leq \mathrm{PSL}_2(\mathbb{R})$ is a Fuchsian group (Definition 34.7.3).

35.1.2. Our first order of business is to try to understand the structure of the classical modular group in terms of this action. Let

$$S := \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}, \quad T := \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix} \in \mathrm{PSL}_2(\mathbb{Z}).$$

Then $Sz = -1/z$ for $z \in \mathbf{H}^2$, so S maps the unit circle $\{z \in \mathbb{C} : |z| = 1\}$ to itself, fixing the point $z = i$; and $Tz = z + 1$ for $z \in \mathbf{H}^2$ acts by translation. We compute that $S^2 = 1$

© The Author(s) 2021

J. Voight, *Quaternion Algebras*, Graduate Texts in Mathematics 288,
https://doi.org/10.1007/978-3-030-56694-4_35

(in $\mathrm{PSL}_2(\mathbb{Z})$) and

$$ST = \begin{pmatrix} 0 & -1 \\ 1 & 1 \end{pmatrix}$$

so $(ST)^3 = 1$.

35.1.3. In a moment, we will see that $\mathrm{PSL}_2(\mathbb{Z})$ is generated by S and T, with a minimal set of relations given by $S^2 = (ST)^3 = 1$. To do so, we examine a fundamental set (cf. Definition 34.1.14) for the action of $\mathrm{PSL}_2(\mathbb{Z})$ on $\mathbf{H}^2$, as follows. Let

$$⌑ := \{z \in \mathbf{H}^2 : |\mathrm{Re}\, z| \le 1/2 \text{ and } |z| \ge 1\}.$$

The set ⌑ is a hyperbolic triangle with vertices at $\omega = (-1+\sqrt{-3})/2$ and $-\omega^2 = (1+\sqrt{-3})/2$ and ∞. The translates of ⌑ by words in S, T tessellate the plane as in Figure 35.1.4.

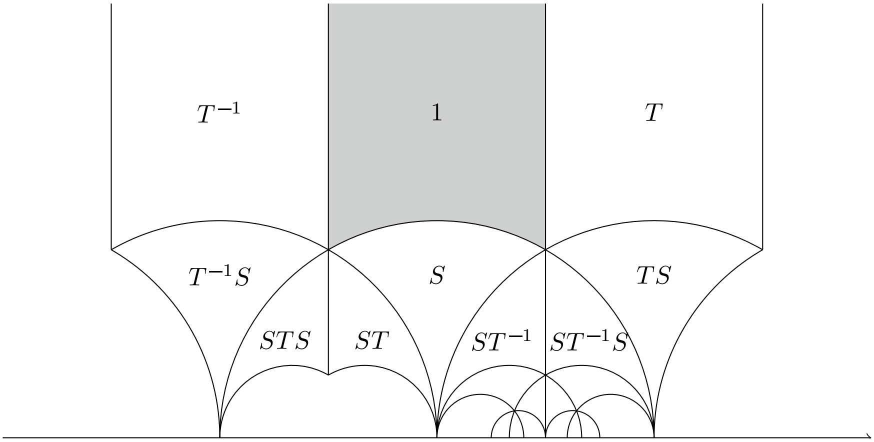


Figure 35.1.4: Tessellation of $\mathbf{H}^2$ into fundamental domains for $\mathrm{SL}_2(\mathbb{Z})$

By the Gauss–Bonnet formula 33.6.8 (or Exercise 35.1),

$$\mathrm{area}(⌑) = \pi - 2\frac{\pi}{3} = \frac{\pi}{3}. \tag{35.1.5}$$

The elements S, T act on the edges of this triangle as in Figure 35.1.6.

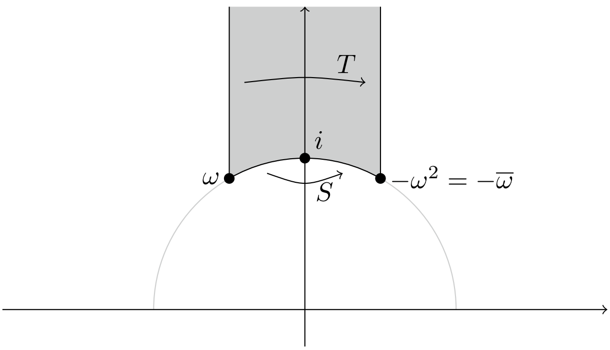


Figure 35.1.6: Action of S, T on ⌑

In the unit disc, the triangle ⌑ is as in Figure 35.1.7.

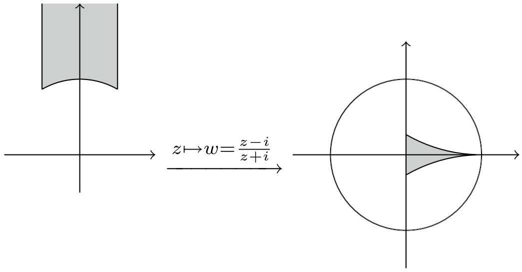


Figure 35.1.7: ⌑ in $\mathbf{D}^2$

The following three lemmas describe the relationship of the set ⌑ to Γ.

Lemma 35.1.8. *For all $z \in \mathbf{H}^2$, there exists a word $\gamma \in \langle S, T \rangle$ such that $\gamma z \in$ ⌑.*

Proof. In fact, we can determine such a word algorithmically. First, we translate z so that $|\operatorname{Re} z| \leq 1/2$. If $|z| \geq 1$, we are done; otherwise, if $|z| < 1$, then

$$\operatorname{Im}\left(\frac{-1}{z}\right) = \frac{\operatorname{Im} z}{|z|^2} > \operatorname{Im} z. \tag{35.1.9}$$

We then repeat this process, obtaining a sequence of elements $z = z_1, z_2, \dots$ with $\operatorname{Im} z_1 < \operatorname{Im} z_2 < \dots$. We claim that this process terminates after finitely many steps. Indeed, by (33.3.9)

$$\operatorname{Im}(gz) = \frac{\operatorname{Im} z}{|cz+d|^2}, \text{ for } g = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in \operatorname{PSL}_2(\mathbb{R}),$$

and the number of $c, d \in \mathbb{Z}$ such that $|cz + d| < 1$ is finite: the set $\mathbb{Z} + \mathbb{Z}z \subseteq \mathbb{C}$ is a lattice, so there are only finitely many elements of bounded norm. (Alternatively, the orbit Γz is discrete by Theorem 34.5.1—or the direct argument given in Proposition 34.7.2—therefore, its intersection with the compact set

$$K = \{z' \in \mathbf{H}^2 : |\operatorname{Re}(z')| \leq 1/2 \text{ and } \operatorname{Im} z \leq \operatorname{Im} z' \leq 1\}$$

is finite.) Upon termination, we have found a word γ in S, T such that $\gamma z \in$ ⌑. □

The procedure exhibited in the proof of Lemma 35.1.8 is called a **reduction algorithm**.

Lemma 35.1.10. *Let $z, z' \in$ ⌑, and suppose $z \in \operatorname{int}($⌑$)$ lies in the interior of ⌑. If $z' = \gamma z$ with $\gamma \in \Gamma$, then $\gamma = 1$ and $z = z'$.*

Proof. Let $z' = \gamma z$ with $\gamma = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in \Gamma$. We have $\operatorname{Im} z' = (\operatorname{Im} z)/|cz+d|^2$. First suppose that $\operatorname{Im} z' \geq \operatorname{Im} z$; then

$$|cz+d|^2 = (c \operatorname{Re} z + d)^2 + c^2 (\operatorname{Im} z)^2 \leq 1. \tag{35.1.11}$$

Since $\operatorname{Im} z > \operatorname{Im}\omega = \sqrt{3}/2$, from (35.1.11) we conclude that $c^2 \leq 4/3$ so $|c| \leq 1$. If $c = 0$ then $ad - bc = ad = 1$ so $a = d = \pm 1$, and then $z' = \gamma z = z \pm b$, which immediately implies $b = 0$ so $\gamma = 1$ as claimed. If instead $|c| = 1$, then the conditions

$$(c \operatorname{Re} z + d)^2 \leq 1 - (\operatorname{Im} z)^2 \leq 1 - 3/4 = 1/4 \quad \textit{and} \quad |\operatorname{Re} z| < 1/2$$

together imply $d = 0$; but then $|cz + d| = |z| \leq 1$, and since $z \in \operatorname{int}(⌑)$ we have $|z| > 1$, a contradiction.

If instead $\operatorname{Im} z' < \operatorname{Im} z$, we interchange the roles of z, z' and have strict inequality in (35.1.11); by the same argument and the weaker inequality $|\operatorname{Re} z| \leq 1/2$, we then obtain $|z| < 1$, a contradiction. □

Lemma 35.1.12. *The elements S, T generate $\Gamma = \mathrm{PSL}_2(\mathbb{Z})$.*

Proof. Let $z = 2i \in \operatorname{int}(⌑)$. Let $\gamma \in \Gamma$, and let $z' = \gamma z$. By Lemma 35.1.8, there exists γ' a word in S, T such that $\gamma' z' \in ⌑$. By Lemma 35.1.10, we have $\gamma' z' = (\gamma'\gamma) z = z$, so $\gamma'\gamma = 1$ and $\gamma = \gamma' \in \langle S, T\rangle$. □

Although we have worked in $\mathrm{PSL}_2(\mathbb{Z})$ throughout, it follows from Lemma 35.1.12 that the matrices S, T also generate $\mathrm{SL}_2(\mathbb{Z})$, since $S^2 = -1$. See Exercise 35.3 for another proof of Lemma 35.1.12.

Corollary 35.1.13. *The set ⌑ is a fundamental set for* $\mathrm{PSL}_2(\mathbb{Z}) \circlearrowright \mathbf{H}^2$.

Proof. The statement follows from Lemmas 35.1.8 and 35.1.10 (recalling the definition of fundamental set, Definition 34.1.14). □

35.1.14. If $z \in ⌑$ has $\operatorname{Stab}_\Gamma(z) \neq \{1\}$, then we claim that one of the following holds:

(i) $z = i$, and $\operatorname{Stab}_\Gamma(i) = \langle S\rangle \simeq \mathbb{Z}/2\mathbb{Z}$;

(ii) $z = \omega$, and $\operatorname{Stab}_\Gamma(\omega) = \langle ST\rangle \simeq \mathbb{Z}/3\mathbb{Z}$; or

(iii) $z = -\omega^2$, and $\operatorname{Stab}_\Gamma(-\omega^2) = \langle TS\rangle = T \operatorname{Stab}_\Gamma(\omega) T^{-1}$.

Indeed, let $\gamma z = z$ with $\gamma = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$ and $\gamma \neq 1$. Then $cz^2 + (d-a)z - b = 0$, so $c \neq 0$ and

$$z = \frac{(a-d) + \sqrt{D}}{2c} \quad \text{where } D = \operatorname{Tr}(\gamma)^2 - 4 \in \mathbb{Z}_{<0}.$$

Thus $D = -4$ or $D = -3$. In either case, since $z \in ⌑$ we have $\operatorname{Im} z \geq \sqrt{3}/2$, we must have $c = \pm 1$, and replacing $\gamma \leftarrow -\gamma$ we may take $c = 1$. If $D = -4$, then $\operatorname{Tr}(\gamma) = a + d = 0$ so $z = a + i$, and $a = 0 = d$ and $c = 1 = -b$, i.e., $z = i$ and we are in case (i). If the discriminant is -3, then a similar argument gives $z = ((a \pm 1) + \sqrt{-3})/2$ so $a = 0$, and we are in cases (ii) or (iii).

Therefore, if $\gamma \in \mathrm{PSL}_2(\mathbb{Z})$ has finite order, then γ fixes a point, thus a conjugate of γ would fix a point in ⌑, and therefore by the above γ is conjugate in $\mathrm{PSL}_2(\mathbb{Z})$ to either S or ST.

Let $Y = \Gamma\backslash\mathbf{H}^2$. Gluing together the fundamental set, we obtain a homeomorphism

$$Y = \Gamma\backslash\mathbf{H}^2 \simeq \mathbb{P}^1(\mathbb{C})\smallsetminus\{\infty\} \simeq \mathbb{C}.$$

The orbit of the limit point ∞ under Γ is $\mathbb{P}^1(\mathbb{Q}) \subseteq \operatorname{bd}\mathbf{H}^2$; letting $\mathbf{H}^{2*} = \mathbf{H}^2 \cup \mathbb{P}^1(\mathbb{Q})$, there is a homeomorphism

$$X = \Gamma\backslash\mathbf{H}^{2*} \simeq \mathbb{P}^1(\mathbb{C})$$

as in Figure 35.1.15.

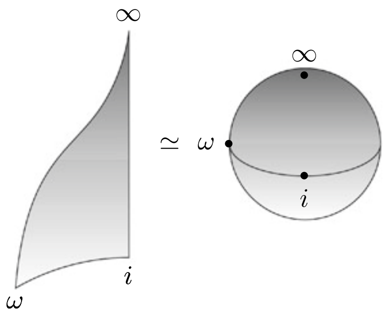


Figure 35.1.15: Y as orbifold and as Riemann surface

Away from the orbits $\Gamma i, \Gamma\omega$ with nontrivial stabilizer, the complex structure on $\mathbf{H}^2$ descends and gives the quotient $Y\smallsetminus\{\Gamma i, \Gamma\omega\}$ the structure of a Riemann surface. By studying the moduli of lattices, later we will give an explicit holomorphic identification $j : Y \to \mathbb{C}$.

35.1.16. By 34.8.11, the quotient Y has the structure of a good complex 1-orbifold, when we keep track of the two nontrivial stabilizers.

Alternatively, we can also give X the structure of a compact Riemann surface as follows. Let $z_0 \in$ ⌑ $\cup \{\infty\}$. If $z_0 = \infty$, we take the chart $z \mapsto e^{2\pi i z}$. Otherwise, let $e = \#\operatorname{Stab}_\Gamma(z_0) < \infty$, let $w = (z - z_0)/(z - \overline{z_0})$ be the local coordinate as in (33.7.3), and take the chart $z \mapsto w^e$ at z_0.

Lemma 35.1.17. *Every relation among S, T is obtained from $S^2 = (ST)^3 = 1$ after conjugation by Γ, so that Γ has the presentation*

$$\Gamma \simeq \langle S, T \mid S^2 = (ST)^3 = 1\rangle.$$

Thus $\Gamma = \mathrm{PSL}_2(\mathbb{Z})$ is the *free product* of $\mathbb{Z}/2\mathbb{Z}$ and $\mathbb{Z}/3\mathbb{Z}$.

Proof. Consider a relation $\delta_1 \cdots \delta_r = 1$ where $\delta_i \in \{S, T, T^{-1}\}$ (recall $S = S^{-1}$). We may suppose that $\delta_i \neq \delta_{i+1}^{-1}$ for $i = 1, \dots, r-1$ or else we can cancel adjacent terms. Define $\gamma_i := \delta_1 \cdots \delta_i$ for $i = 1, \dots, r$, so that $\gamma_r = 1$ and $\gamma_{i+1} = \gamma_i\delta_{i+1}$ for all $i = 1, \dots, r-1$. Let $z_0 = 2i$ and let

$$z_i := \gamma_i z_0 = (\delta_1 \cdots \delta_i) z_0.$$

We claim that z_{i+1} and z_i are in adjacent Γ-translates of ⌑: indeed,

$$\gamma_{i+1}⌑ \cap \gamma_i ⌑ \in \gamma_i(\delta_{i+1}⌑ \cap ⌑) \neq \emptyset \tag{35.1.18}$$

is a side of both. Draw the geodesic (shortest path, across the corresponding side) between z_{i+1} and z_i for each i. Because $\delta_i \neq \delta_{i+1}^{-1}$, there is no backtracking; taken together, they define a loop in the upper half-plane, since $z_r = \gamma_r z_0 = z_0$.

We first conjugate the relation by T or T^{-1} so $\delta_1 = S$. If $\delta_r = S$ as well, then we conjugate by S and begin again. So without loss of generality $\delta_r = T, T^{-1}$; we explain the case $\delta_r = T$, the case $\delta_r = T^{-1}$ is similar. Then our relation looks like $S\delta_2 \cdots \delta_{r-1}T = 1$, and so $\gamma_r = 1 = \gamma_{r-1}\delta_r = \gamma_{r-1}T$ implies $z_{r-1} = \gamma_{r-1}z_0 = T^{-1}z_0 = 2i - 1$.

If $z_i = \gamma_i z_0 = z_0$ for some $0 < i < r$, then $\gamma_i = \delta_1 \cdots \delta_i = 1$ and similarly $\delta_{i+1} \cdots \delta_r = 1$, so we may argue separately with each such relation and so may suppose that $z_i \neq z_0$ for all $i = 1, \dots, r-1$. It follows that the loop intersects ⌑ only in the path from z_0 to z_1 and from z_{r-1} to z_0, because any other intersection would necessarily have source or target z_0.

We observe that ω is in the interior of the loop, since by continuity any path from i to $2i - 1/2$ in $\mathbf{H}^2$ that does not intersect ⌑ must go from right to left through a highest value $-1/2 + it$ with $0 < t \leq \sqrt{3}/2$.

The proof proceeds by induction on the number of points in the intersection of the interior of the loop with the set of vertices $\Gamma\omega$. We have shown in the previous paragraph that if there are no such points, then the relation is trivial. In the general case with relation $S\delta_2 \cdots \delta_{r-1}T = 1$, expanding $(ST)^3 = 1$ we get $T = (ST^{-1})^2 S$ and substituting we obtain another relation $S\delta_2 \cdots \delta_{r-1}(ST^{-1})^2 S = 1$, as in Figure 35.1.19.

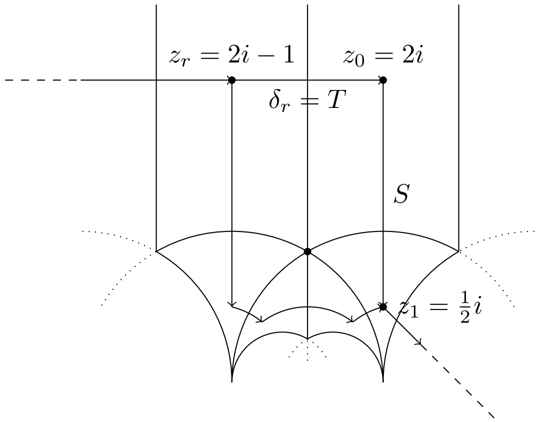


Figure 35.1.19: Simplifying a relation

Reading this relation as above, we see that the loop encloses one fewer point in $\Gamma\omega$ (it starts and ends with a backtracking step), and so the same is true for the conjugate relation $\delta_2 \cdots \delta_{r-1}(ST^{-1})^2 = 1$. Cancelling any new adjacent terms and conjugating the relation does not change the number of enclosed interior vertices, so the result holds by induction.

(Alternatively, for a proof in the style of Lemma 35.1.10, see Exercise 35.5.) □

Remark 35.1.20. Alperin [Alp93] uses the action on the irrational numbers to show directly that $\mathrm{PSL}_2(\mathbb{Z})$ is the free product of $\mathbb{Z}/2\mathbb{Z}$ and $\mathbb{Z}/3\mathbb{Z}$ (but note the typo $\beta(z) = 1 - 1/z$ on the first page).

35.2 Binary quadratic forms

We pause to give an application to quadratic forms and class groups, after Gauss. An **integral binary quadratic form**, abbreviated in this section to simply **form**, is an expression

$$Q(x, y) = ax^2 + bxy + cy^2 \in \mathbb{Z}[x, y].$$

We define the **discriminant** of a form Q to be

$$\mathrm{disc}(Q) := b^2 - 4ac.$$

A form Q is **primitive** if $\gcd(a, b, c) = 1$, and Q is **positive definite** if $Q(x, y) > 0$ for all nonzero $(x, y) \in \mathbb{R}^2$; after completing the square, we see that a form is positive definite if and only if $a > 0$ and $\mathrm{disc}(Q) < 0$.

For $d < 0$, let

$$\mathcal{Q}_d := \{Q(x, y) = ax^2 + bxy + cy^2 : a > 0,\ \mathrm{disc}(Q) = d\}$$

be the set of primitive, positive definite forms of discriminant d. The group Γ acts on $\mathcal{Q}_d$ the *right* by change of variable: for $\gamma \in \Gamma$, we define $(Q^\gamma)(x, y) = Q(\gamma(x, y)^{\mathsf{t}})$, so that if $\gamma = \begin{pmatrix} r & s \\ t & u \end{pmatrix}$ then

$$(Q^\gamma)(x, y) = Q(rx + sy, tx + uy).$$

We verify that $\mathrm{disc}(Q^\gamma) = \mathrm{disc}(Q) = d$ for $\gamma \in \Gamma$. We say that Q, Q' are (Γ-)**equivalent** if $Q' = Q^\gamma$ for some $\gamma \in \Gamma = \mathrm{SL}_2(\mathbb{Z})$.

We claim that the number of equivalence classes $h(d) := \#\mathcal{Q}_d/\Gamma$ is finite. Indeed, to every $Q \in \mathcal{Q}_d$, we associate the unique root

$$z_Q = \frac{-b + \sqrt{|d|}i}{2a} \in \mathbf{H}^2$$

of $Q(z, 1) = 0$. Then $z_{Q^\gamma} = \gamma^{-1}(z)$ for $\gamma \in \Gamma$. Therefore, by the reduction theory of the previous section, we can replace Q up to equivalence by a form such that $z_Q \in ⌑$. If we further insist that $\mathrm{Re}\, z < 1/2$ and $\mathrm{Re}\, z < 0$ if $|z| = 1$, then this representative is

unique, as in Figure 35.2.1.

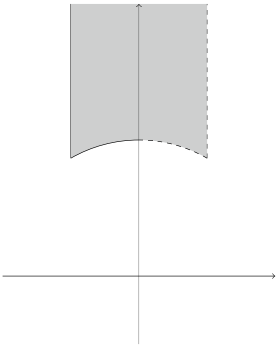


Figure 35.2.1: A choice of unique representative in ⌑

Thus
$$-\frac{1}{2} \leq \operatorname{Re} z_Q = -\frac{b}{2a} < \frac{1}{2}$$
so $-a < b \leq a$, or equivalently,
$$|b| \leq a \ and \ (b \geq 0 \ if \ |b| = a);$$
and
$$|z_Q| = \frac{b^2 - d}{4a^2} = \frac{c}{a} \geq 1$$
so $a \leq c$ and $b \geq 0$ if equality holds. In sum, every positive definite form Q is equivalent to a ($\mathrm{SL}_2(\mathbb{Z})$-)**reduced** form satisfying
$$|b| \leq a \leq c \qquad \text{with } b \geq 0 \text{ if } |b| = a \text{ or } a = c.$$

We now show that there are only finitely many reduced forms with given discriminant $d < 0$, i.e., that $h(d) < \infty$. The inequalities $|b| \leq a \leq c$ imply that
$$|d| = 4ac - b^2 \geq 3a^2,$$
so $a \leq \sqrt{|d|/3}$ and $|b| \leq a$, so there are only finitely many possibilities for a, b; and then $c = (b^2 - d)/(4a)$ is determined. This gives an efficient method to compute the set $\mathcal{Q}_d/\Gamma$ efficiently.

Let $S = \mathbb{Z} \oplus \mathbb{Z}[(d + \sqrt{d})/2] \subset K = \mathbb{Q}(\sqrt{d})$ be the quadratic ring of discriminant $d < 0$. Let $\operatorname{Pic}(S)$ be the group of invertible fractional ideals of S modulo principal ideals. Then there is a bijection
$$\mathcal{Q}_d/\Gamma \leftrightarrow \operatorname{Pic}(S)$$
$$[ax^2 + bxy + cy^2] \mapsto [\mathfrak{a}] = \left[\left(a, \frac{-b + \sqrt{d}}{2}\right)\right]$$
(Exercise 35.9). In the same stroke, we have proven the finiteness of the class number $\#\operatorname{Pic}(S) < \infty$.

35.3 Moduli of lattices

In this section, we realize $\mathrm{PSL}_2(\mathbb{Z})\backslash\mathbf{H}^2$ as a moduli space of complex lattices.

35.3.1. A **(complex) lattice** $\Lambda \subset \mathbb{C}$ is a subgroup $\Lambda = \mathbb{Z}z_1 + \mathbb{Z}z_2$ with z_1, z_2 linearly independent over $\mathbb{R}$; the elements z_1, z_2 are a basis for Λ.

Two lattices Λ, Λ' are **homothetic** if there exists $u \in \mathbb{C}^\times$ such that $\Lambda' = u\Lambda$, and we write $\Lambda \sim \Lambda'$. Let $\Lambda = \mathbb{Z}z_1 + \mathbb{Z}z_2$ be a lattice. Then without loss of generality (interchanging z_1, z_2), we may suppose $\mathrm{Im}(z_2/z_1) > 0$, and then we call z_1, z_2 an **oriented basis**. Then there is a homothety

$$\Lambda \sim \frac{1}{z_1}\Lambda = \mathbb{Z} + \mathbb{Z}\tau$$

where $\tau = z_2/z_1 \in \mathbf{H}^2$.

Lemma 35.3.2. *Let $\Lambda = \mathbb{Z} + \mathbb{Z}\tau$ and $\Lambda = \mathbb{Z} + \mathbb{Z}\tau'$ be lattices with $\tau, \tau' \in \mathbf{H}^2$. Then $\Lambda \sim \Lambda'$ if and only if $\Gamma\tau = \Gamma\tau'$.*

Proof. Since $\tau, \tau' \in \mathbf{H}^2$, the bases $1, \tau$ and $1, \tau'$ are oriented. We have $\Lambda = \mathbb{Z} + \mathbb{Z}\tau \sim \mathbb{Z} + \mathbb{Z}\tau' = \Lambda'$ if and only if there exists an invertible change of basis matrix $g = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in \mathrm{GL}_2(\mathbb{Z})$ and $u \in \mathbb{C}^\times$ such that

$$u\begin{pmatrix} a & b \\ c & d \end{pmatrix}\begin{pmatrix} \tau \\ 1 \end{pmatrix} = \begin{pmatrix} \tau' \\ 1 \end{pmatrix}$$

so $u(a\tau + b) = \tau'$ and $u(c\tau + d) = 1$. Eliminating u gives equivalently

$$\tau' = \frac{a\tau + b}{c\tau + d}.$$

Therefore $g \in \mathrm{SL}_2(\mathbb{Z})$, and since g is well-defined as an element of $\Gamma = \mathrm{PSL}_2(\mathbb{Z})$, the result follows. □

35.3.3. By Lemma 35.3.2, there is a bijection

$$\begin{aligned} Y = \Gamma\backslash\mathbf{H}^2 &\to \{\Lambda \subset \mathbb{C} \text{ lattice}\}/\sim \\ \Gamma\tau &\mapsto [\mathbb{Z} + \mathbb{Z}\tau]; \end{aligned} \tag{35.3.4}$$

that is to say, $Y = \Gamma\backslash\mathbf{H}^2$ parametrizes complex lattices up to homothety.

To a lattice Λ, we associate the complex torus $\mathbb{C}/\Lambda$ (of rank 1); two such tori $\mathbb{C}/\Lambda$ and $\mathbb{C}/\Lambda'$ are isomorphic as Riemann surfaces if and only if $\Lambda \sim \Lambda'$. Therefore, the space Y also parametrizes complex tori.

We return to this interpretation in section 40.1.

35.4 Congruence subgroups

The finite-index subgroups of $\mathrm{PSL}_2(\mathbb{Z})$ play a central role, and of particular importance are those subgroups defined by congruence conditions on the entries.

Definition 35.4.1. Let $N \in \mathbb{Z}_{\geq 1}$. The **full congruence subgroup** $\Gamma(N) \trianglelefteq \mathrm{PSL}_2(\mathbb{Z})$ of **level** N is

$$\begin{aligned}\Gamma(N) &:= \ker(\mathrm{PSL}_2(\mathbb{Z}) \to \mathrm{PSL}_2(\mathbb{Z}/N\mathbb{Z})) \\ &= \left\{\gamma \in \mathrm{PSL}_2(\mathbb{Z}) : \gamma = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \equiv \pm\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \pmod N\right\}.\end{aligned}$$

To avoid confusion, from now on we will now write $\Gamma(1) = \mathrm{PSL}_2(\mathbb{Z})$.

35.4.2. By strong approximation for $\mathrm{SL}_2(\mathbb{Z})$ (Theorem 28.2.6), the map $\mathrm{SL}_2(\mathbb{Z}) \to \mathrm{SL}_2(\mathbb{Z}/N\mathbb{Z})$ is surjective for all $N \geq 1$, so there is an exact sequence

$$1 \to \Gamma(N) \to \Gamma(1) \to \mathrm{PSL}_2(\mathbb{Z}/N\mathbb{Z}) \to 1.$$

Definition 35.4.3. A subgroup $\Gamma \leq \Gamma(1)$ is a **congruence subgroup** if $\Gamma \geq \Gamma(N)$ for some $N \geq 1$; if so, the minimal such N is called the **level** of Γ.

Remark 35.4.4. Noncongruence subgroups (finite-index subgroups not containing $\Gamma(N)$ for any $N \geq 1$) also play a role in the structure of the group $\mathrm{SL}_2(\mathbb{Z})$: see the recent survey by Li–Long [LL2012] and the references therein.

35.4.5. In addition to the congruence groups $\Gamma(N)$ themselves, we will make use of two other important congruence subgroups for $N \geq 1$:

$$\begin{aligned}\Gamma_0(N) &:= \left\{\gamma = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in \mathrm{PSL}_2(\mathbb{Z}) : c \equiv 0 \pmod N\right\} \\ &= \left\{\gamma \in \mathrm{PSL}_2(\mathbb{Z}) : \gamma \equiv \pm\begin{pmatrix} * & * \\ 0 & * \end{pmatrix} \pmod N\right\} \\ \Gamma_1(N) &:= \left\{\gamma \in \mathrm{PSL}_2(\mathbb{Z}) : \gamma \equiv \pm\begin{pmatrix} 1 & * \\ 0 & 1 \end{pmatrix} \pmod N\right\}\end{aligned} \tag{35.4.6}$$

Visibly, $\Gamma(N) \leq \Gamma_1(N) \leq \Gamma_0(N)$. We accordingly write

$$\begin{aligned}Y_0(N) &:= \Gamma_0(N)\backslash\mathbf{H}^2 \\ X_0(N) &:= \Gamma_0(N)\backslash\mathbf{H}^{2*}\end{aligned} \tag{35.4.7}$$

where $\mathbf{H}^{2*} := \mathbf{H}^2 \cup \mathbb{P}^1(\mathbb{Q})$, and similarly $Y_1(N)$ and $Y(N)$.

In the remainder of this section, we consider as an extended example the case $N = 2$. We can equally well write

$$\Gamma(2) = \left\{\pm\begin{pmatrix} a & b \\ c & d \end{pmatrix} \in \mathrm{PSL}_2(\mathbb{Z}) : b \equiv c \equiv 0 \pmod 2\right\} \tag{35.4.8}$$

From 35.4.2,

$$\Gamma(1)/\Gamma(2) \simeq \mathrm{PSL}_2(\mathbb{Z}/2\mathbb{Z}) = \mathrm{GL}_2(\mathbb{F}_2) \simeq S_3 \tag{35.4.9}$$

the nonabelian group of order 6, so in particular $[\Gamma(1) : \Gamma(2)] = 6$.

We can uncover the structure of the group $\Gamma(2)$ in a manner similar to what we did for $\Gamma(1)$ in section 35.1—the details are requested in Exercise 35.10. The group $\Gamma(2)$ is generated by

$$\begin{pmatrix} 1 & 2 \\ 0 & 1 \end{pmatrix} \text{ and } \begin{pmatrix} 1 & 0 \\ 2 & 1 \end{pmatrix}$$

which act on $\mathbf{H}^2$ by $z \mapsto z + 2$ and $z \mapsto z/(2z + 1)$, respectively, and a fundamental set $\mathrel{\square}$ is given in Figure 35.4.10.

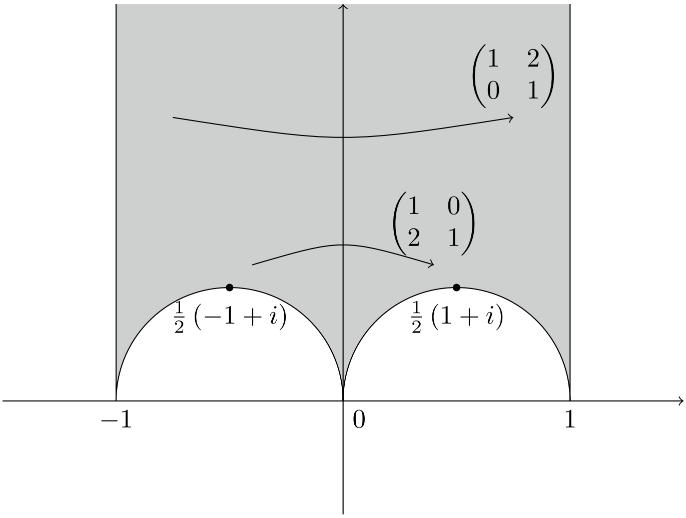


Figure 35.4.10: A fundamental set for $\Gamma(2) \circlearrowright \mathbf{H}^2$

(In fact, later we will see from more general structural results that $\Gamma(2)$ is freely generated by these two elements, so it is isomorphic to the free group on two generators.)

The action $\Gamma(2) \circlearrowright \mathbf{H}^2$ is free: by 35.1.14, if $\gamma z = z$ with $z \in \mathbf{H}^2$ and $\gamma \in \Gamma(2) \leq \Gamma(1)$, then γ is conjugate in $\Gamma(1)$ to either S, ST; but $\Gamma(2) \trianglelefteq \Gamma(1)$ is normal, so without loss of generality either $S = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}$ or $ST = \begin{pmatrix} 0 & -1 \\ 1 & 1 \end{pmatrix}$ belongs to $\Gamma(2)$, a contradiction.

Let $Y(2) := \Gamma(2)\backslash\mathbf{H}^2$. Then gluing together the fundamental set, there is a homeomorphism

$$Y(2) \simeq \mathbb{P}^1(\mathbb{C}) \smallsetminus \{0, 1, \infty\}.$$

The limit points of $\mathrel{\square}$ in $\operatorname{bd}\mathbf{H}^2$ are the points $-1, 0, 1, \infty$ and the points $-1, 1$ are identified in the quotient (by translation). The orbit of these points under $\Gamma(2)$ is $\mathbb{P}^1(\mathbb{Q}) \subseteq \operatorname{bd}\mathbf{H}^2$, so letting $\mathbf{H}^{2*} = \mathbf{H}^2 \cup \mathbb{P}^1(\mathbb{Q})$, there is a homeomorphism

$$X(2) := \Gamma(2)\backslash\mathcal{H}^* \simeq \mathbb{P}^1(\mathbb{C}). \tag{35.4.11}$$

We have a natural holomorphic projection map

$$X(2) = \Gamma(2)\backslash \mathbf{H}^{2*} \to X(1) = \Gamma(1)\backslash \mathbf{H}^{2*}; \qquad (35.4.12)$$

via (35.4.9), the group $\mathrm{GL}_2(\mathbb{F}_2)$ acts on $X(2)$ by automorphisms:

$$\begin{aligned} \mathrm{GL}_2(\mathbb{F}_2) \circlearrowright X(2) &\to X(2) \\ \gamma(\Gamma(2)z) &= \Gamma(2)\gamma z \end{aligned}$$

where $\gamma \in \Gamma(1)$ is a lift, so the map (35.4.12) is obtained as the quotient by $\mathrm{GL}_2(\mathbb{F}_2)$.

Finally, the congruence conditions (35.4.6) imply that $\Gamma_0(2) = \Gamma_1(2)$ has index 2 in $\Gamma(2)$, with the quotient generated by T, and we obtain a fundamental set by identifying the two ideal triangles in ⌑ above.

Exercises

1. Prove that $Y(1) = \mathrm{SL}_2(\mathbb{Z})\backslash \mathbf{H}^2$ has $\operatorname{area}(Y(1)) = \pi/3$ by direct integration (verifying the Gauss–Bonnet formula).

2. Show that $\mathrm{PSL}_2(\mathbb{Z})$ is generated by T and $U = \begin{pmatrix} 1 & 0 \\ 1 & 1 \end{pmatrix}$. [So $\mathrm{PSL}_2(\mathbb{Z})$ is generated by two parabolic elements (of infinite order), just as it is generated by elements of order two and three.]

3. Prove Lemma 35.1.12 using Lemma 28.3.3 (elementary matrices).

4. In this exercise, we link the fact that $\mathrm{PSL}_2(\mathbb{Z})$ is generated by S, T to a kind of continued fraction via the Euclidean algorithm. [So the reduction algorithm is a way to visualize the Euclidean algorithm.] Let $a, b \in \mathbb{Z}_{\geq 1}$ with $a \geq b$.

   (a) Show that there exist unique $q, r \in \mathbb{Z}$ such that $a = qb - r$ and $q \geq 2$ and $0 \leq r < b$.

   From (a), define inductively $r_0 = a$, $r_1 = b$, and $r_{i-1} = q_i r_i - r_{i+1}$ with $0 \leq r_{i+1} < r_i$; we then have $r_1 > r_2 > \cdots > r_t > r_{t+1} = 0$ for some $t > 0$.

   b) Show that $\gcd(a, b) = r_t$, and if $\gcd(a, b) = 1$ then

   $$\frac{a}{b} = q_1 - \cfrac{1}{q_2 - \cfrac{1}{\cdots - \cfrac{1}{q_t}}}.$$

   Such a continued fraction is called a **negative-regular** or **Hirzebruch–Jung continued fraction**. [The Hirzebruch–Jung continued fraction plays a role in the resolution of singularities [Jun08, Hir53].]

   c) Show (by induction) that

   $$\begin{pmatrix} 0 & 1 \\ -1 & q_t \end{pmatrix} \cdots \begin{pmatrix} 0 & 1 \\ -1 & q_1 \end{pmatrix} \begin{pmatrix} a \\ b \end{pmatrix} = \begin{pmatrix} r_t \\ 0 \end{pmatrix}.$$

For all $q \in \mathbb{Z}$, write

$$\begin{pmatrix} 0 & 1 \\ -1 & q \end{pmatrix} \in \langle S, T \rangle \subseteq \mathrm{PSL}_2(\mathbb{Z})$$

as a word in S, T, and interpret the action of this matrix in terms of the reduction algorithm to the fundamental set ⌑ for $\mathrm{PSL}_2(\mathbb{Z})$.

d) Let $A = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in \mathrm{SL}_2(\mathbb{Z})$. Show that $\gcd(a, c) = 1$ and conclude from (c) that there exists $W \in \langle S, T \rangle$ such that

$$WA = \begin{pmatrix} 1 & b' \\ 0 & 1 \end{pmatrix}$$

with $b' \in \mathbb{Z}$. Conclude that $\langle S, T \rangle = \mathrm{PSL}_2(\mathbb{Z})$. (So how, in the end, does this procedure to write A in terms of S and T relate to the one given by the reduction algorithm in Lemma 35.1.8?)

5. In this exercise, we give a "matrix proof" that a complete set of relations satisfied by S, T in $\mathrm{PSL}_2(\mathbb{Z})$ are $S^2 = (ST)^3 = 1$.
   (a) Show that it suffices to show that no word $S(ST)^{e_1} S(ST)^{e_2} \dots S(ST)^{e_n}$ with $e_i = 1, 2$ is equal to 1.
   (b) Observe that $S(ST) = T$ and $S(ST)^2$ have at least one off-diagonal entry nonzero and can be represented with a matrix whose entries all have the same sign.
   (c) Show that if $A = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$ has at least one off-diagonal entry nonzero and all entries of the same sign, then these properties hold also for both $S(ST)A$ and $S(ST)^2 A$. Conclude that (a) holds.

   [This argument is given by Fine [Fin89, Theorem 3.2.1].]
6. Show that the commutator subgroup $\Gamma' \trianglelefteq \Gamma = \mathrm{PSL}_2(\mathbb{Z})$ (the subgroup generated by commutators $\gamma\delta\gamma^{-1}\delta^{-1}$ for $\gamma, \delta \in \Gamma$) has index 6 and $\Gamma/\Gamma' \simeq \mathbb{Z}/6\mathbb{Z}$.
7. Compute the class number $h(d)$ and the set of reduced (positive definite) binary quadratic forms of discriminant $d = -71$.
8. Let $\mathcal{Q}_d$ be the set of primitive, positive definite binary quadratic forms of discriminant $d < 0$ and let $\mathcal{Q} = \bigcup_d \mathcal{Q}_d$.
   (a) Show that the group $\mathrm{GL}_2(\mathbb{Z})$ acts naturally on $\mathcal{Q}$ by change of variables, with $\mathrm{PGL}_2(\mathbb{Z})$ acting faithfully.
   (b) Consider the action of $\mathrm{PGL}_2(\mathbb{Z})$ on $\mathbf{H}^2$. Show that every $Q \in \mathcal{Q}_d$ is equivalent to a $\mathrm{GL}_2(\mathbb{Z})$-**reduced** form $ax^2 + bxy + cy^2$ satisfying

   $$0 \le b \le a \le c.$$

   [Hint: Find a nice fundamental set ⌑ for $\mathrm{PGL}_2(\mathbb{Z})$.*]*
   (c) By transport, 35.1.14 computes the stabilizer of $\mathrm{PSL}_2(\mathbb{Z})$ on $Q \in \mathcal{Q}_d$. Compute $\mathrm{Stab}_{\mathrm{PGL}_2(\mathbb{Z})}(Q)$ for $Q \in \mathcal{Q}_d$.

▹ 9. Let

$$S = \mathbb{Z} \oplus \mathbb{Z}[(d + \sqrt{d})/2] \subset K = \mathbb{Q}(\sqrt{d})$$

be the quadratic ring of discriminant $d < 0$. Let $\mathrm{Pic}(S)$ be the group of invertible fractional ideals of S modulo principal ideals. Show that the map

$$\mathcal{Q}_d/\Gamma \to \mathrm{Pic}(S)$$
$$[ax^2 + bxy + cy^2] \mapsto [\mathfrak{a}] = \left[\left(a, \frac{-b + \sqrt{d}}{2}\right)\right]$$

is a bijection, where $\mathcal{Q}_d/\Gamma$ is the set of ($\mathrm{SL}_2(\mathbb{Z})$-)equivalence classes of (primitive, positive definite) binary quadratic forms of discriminant d.

▹ 10. Show that the elements

$$\begin{pmatrix} 1 & 2 \\ 0 & 1 \end{pmatrix}, \begin{pmatrix} 1 & 0 \\ 2 & 1 \end{pmatrix}$$

generate $\Gamma(2)$ using the fundamental set

⌑ $= \{z \in \mathbf{H}^2 : |\mathrm{Re}\, z| \leq 1, |2z \pm 1| \geq 1\}.$

[Hint: adapt the method used in section 35.1.]

This chapter is licensed under the terms of the Creative Commons Attribution-NonCommercial 4.0 International License (http://creativecommons.org/licenses/by-nc/4.0/), which permits any noncommercial use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license and indicate if changes were made.

The images or other third party material in this chapter are included in the chapter's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the chapter's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.

Chapter 36

Hyperbolic space

In this chapter, we extend the notions introduced for the hyperbolic plane to hyperbolic space in three dimensions; we follow essentially the same outline, and so our exposition is similarly brief.

36.1 Hyperbolic space

A general, encyclopedic reference for hyperbolic geometry is the book by Ratcliffe [Rat2006]. For further reference, see also Elstrodt–Grunewald–Mennicke [EGM98, Chapter 1], Iversen [Ive92, Chapter VIII], and Marden [Mard2007].

Definition 36.1.1. The **upper half-space** is the set

$$\mathbf{H}^3 := \mathbb{C} \times \mathbb{R}_{>0} = \{(x, y) = (x_1 + x_2 i, y) \in \mathbb{C} \times \mathbb{R} : y > 0\}.$$

Hyperbolic space is the set $\mathbf{H}^3$ equipped with the metric induced by the **hyperbolic length element**

$$\mathrm{d}s^2 := \frac{|\mathrm{d}x|^2 + \mathrm{d}y^2}{y^2} = \frac{\mathrm{d}x_1^2 + \mathrm{d}x_2^2 + \mathrm{d}y^2}{y^2}.$$

36.1.2. The space $\mathbf{H}^3$ is the unique three-dimensional (connected and) simply connected Riemannian manifold with constant sectional curvature -1. The **volume element** corresponding to the hyperbolic length element is accordingly

$$\mathrm{d}V := \frac{\mathrm{d}x_1\,\mathrm{d}x_2\,\mathrm{d}y}{y^3}.$$

36.1.3. A vertical half-plane in hyperbolic space is a set of points with y arbitrary and the coordinate x confined to a line in $\mathbb{C}$. The hyperbolic length element restricted to every vertical half-plane is (equivalent to) the hyperbolic length element on the hyperbolic plane. Therefore, $\mathbf{H}^3$ contains many isometrically embedded copies of $\mathbf{H}^2$.

36.1.4. The **sphere at infinity** is the set

$$\mathrm{bd}\,\mathbf{H}^3 = \mathbb{P}^1(\mathbb{C}) = \mathbb{C} \cup \{\infty\}$$

© The Author(s) 2021

J. Voight, *Quaternion Algebras*, Graduate Texts in Mathematics 288,
https://doi.org/10.1007/978-3-030-56694-4_36

(analogous to the circle at infinity for $\mathbf{H}^2$), with the image of $\mathbb{C}$ corresponding to the locus of points with $t = 0$. We then define the **completed upper half-space** to be

$$\mathbf{H}^{3*} := \mathbf{H}^3 \cup \operatorname{bd} \mathbf{H}^3.$$

The topology on $\mathbf{H}^{3*}$ is defined by taking a fundamental system of neighborhoods of the point at ∞ to be sets of the form

$$\{(x, y) \in \mathbf{H}^3 : y > M\} \cup \{\infty\}$$

for $M > 0$, and the open balls tangent to $z \in \mathbb{C}$ together with z.

36.1.5. The metric space $\mathbf{H}^3$ is complete, and the topology on $\mathbf{H}^3$ is the same as the topology induced by the Euclidean metric.

The geodesics in $\mathbf{H}^3$ are the Euclidean hemicircles orthogonal to $\mathbb{C}$ and vertical half-lines: every two points lie in a vertical hyperbolic plane (see 36.1.3), so this statement can be deduced from the case of the hyperbolic plane. (Alternatively, by applying an element of $\mathrm{PSL}_2(\mathbb{C})$ it is enough to show that the vertical axis $Z = \{(0, y) : y > 0\}$ is a geodesic, and arguing as in (33.5.4) we obtain the result.) Accordingly, $\mathbf{H}^3$ is a uniquely geodesic space.

36.1.6. Just as in distinct points determine a geodesic, so do three distinct points determine a **geodesic plane**, the union of all geodesics through the third point and a point on the geodesic between the other two (the choice taken arbitrarily). In a geodesic plane, the geodesic between two points in the plane is contained in the plane. By the preceding paragraph, the geodesic planes in $\mathbf{H}^3$ are the Euclidean hemispheres orthogonal to $\mathbb{C}$ and the vertical half-planes, as in Figure 36.1.7.

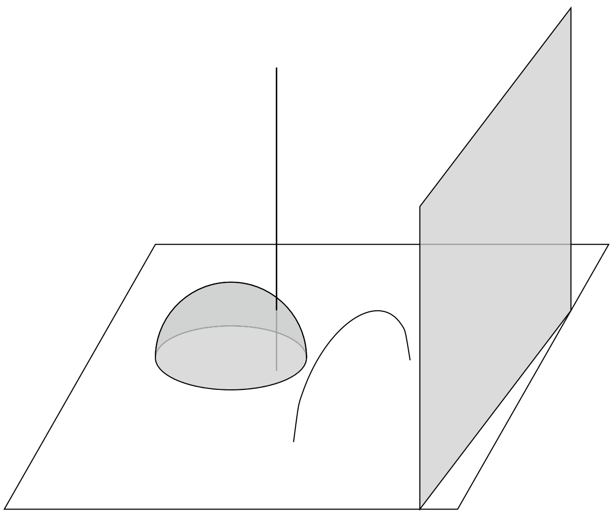


Figure 36.1.7: Geodesic lines and planes in $\mathbf{H}^3$

36.2 Isometries

Analogous to the case of $\mathbf{H}^2$, with orientation-preserving isometries given by $\mathrm{PSL}_2(\mathbb{R})$ acting by linear fractional transformations, in this section we identify the isometries of hyperbolic space $\mathbf{H}^3$ as coming similarly from $\mathrm{PSL}_2(\mathbb{C})$.

36.2.1. The group $\mathrm{PSL}_2(\mathbb{C})$ acts on the sphere at infinity $\mathbb{P}^1(\mathbb{C})$ by linear fractional transformations. We extend this action to $\mathbf{H}^3$ (with almost the same definition!) as follows. We identify

$$\begin{aligned} \mathbf{H}^3 &\hookrightarrow \mathbb{H} = \mathbb{C} + \mathbb{C}j \\ (x, y) &\mapsto z = x + yj \end{aligned}$$

where we recall that $jx = \overline{x}j$ for $x \in \mathbb{C} = \mathbb{R} + \mathbb{R}i \subseteq \mathbb{H}$. We then define the action map

$$\begin{aligned} \mathrm{SL}_2(\mathbb{C}) \times \mathbf{H}^3 &\to \mathbf{H}^3 \\ (g, z) &\mapsto gz = (az + b)(cz + d)^{-1} \end{aligned} \tag{36.2.2}$$

for $g = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in \mathrm{SL}_2(\mathbb{C})$. If $z = x + yj$, then in coordinates (Exercise 36.3)

$$g(z) = \frac{(ax + b)\overline{(cx + d)} + a\overline{c}y^2 + yj}{\|cz + d\|^2} \tag{36.2.3}$$

where

$$\|cz + d\|^2 = \mathrm{nrd}(cz + d) = |cx + d|^2 + |c|^2 y^2.$$

Therefore the image of this map lies in $\mathbf{H}^3$. (Compare this formula with the action of $\mathrm{SL}_2(\mathbb{R})$ in (33.3.8).)

Lemma 36.2.4. *The map (36.2.2) defines a group action of* $\mathrm{SL}_2(\mathbb{C})$ *on* $\mathbf{H}^3$.

Proof. We define the **quaternionic projective line** to be the set

$$\mathbb{P}^1(\mathbb{H}) := \{(\alpha, \beta) : \alpha, \beta \neq (0, 0) \in \mathbb{H}^\times\}/\sim$$

under the equivalence relation $(\alpha, \beta) \sim (\alpha\gamma, \beta\gamma)$ for $\gamma \in \mathbb{H}^\times$, and we denote by $(\alpha : \beta) \in \mathbb{P}^1(\mathbb{H})$ the equivalence class of (α, β). We verify that the group $\mathrm{SL}_2(\mathbb{C})$ acts on $\mathbb{P}^1(\mathbb{H})$ by

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix} \cdot (\alpha : \beta) = (a\alpha + b\beta : c\alpha + d\beta);$$

the left action of $\mathrm{SL}_2(\mathbb{C})$ commutes with the right action of $\mathbb{H}^\times$. The restriction of this action to $\mathbf{H}^3 \hookrightarrow \mathbb{P}^1(\mathbb{H})$ by $z \mapsto (z : 1)$ is

$$(z : 1) \mapsto (az + b : cz + d) = ((az + b)(cz + d)^{-1} : 1)$$

as above. ◻

We now show that $\mathrm{PSL}_2(\mathbb{C})$ acts on $\mathbf{H}^3$ by isometries. This can be verified directly by the formula, with some effort; we prefer to verify this on a convenient set of generators, and so we are first led already to the following decomposition of $\mathrm{SL}_2(\mathbb{C})$ (cf. Proposition 33.4.2).

36.2.5. Let

$$\begin{aligned} K &:= \mathrm{SU}(2) = \left\{ \begin{pmatrix} a & b \\ -\overline{b} & \overline{a} \end{pmatrix} \in \mathrm{M}_2(\mathbb{C}) : |a|^2 + |b|^2 = 1 \right\} \simeq \mathbb{H}^1 \\ A &:= \left\{ \begin{pmatrix} a & 0 \\ 0 & 1/a \end{pmatrix} : a \in \mathbb{R}^{\times}_{>0} \right\} \simeq \mathbb{R} \\ N &:= \left\{ \begin{pmatrix} 1 & b \\ 0 & 1 \end{pmatrix} : b \in \mathbb{C} \right\} \simeq \mathbb{C}. \end{aligned}$$

We have $K = \mathrm{Stab}_{\mathrm{SL}_2(\mathbb{C})}(j)$: from (36.2.3), we see that $gj = (aj+b)(cj+d)^{-1}$ if and only if $|c|^2 + |d|^2 = 1$ and $a\overline{c} + b\overline{d} = 0$; plugging the first equation into the second, and using $ad - bc = 1$ gives $a = \overline{d}$ and then $b = -\overline{c}$.

Letting $z = x + yj$, the other elements act as:

$$\begin{aligned} \begin{pmatrix} a & 0 \\ 0 & 1/a \end{pmatrix}(z) &= a^2(x + yj), \\ \begin{pmatrix} 1 & b \\ 0 & 1 \end{pmatrix}(z) &= (x + b) + yj. \end{aligned} \tag{36.2.6}$$

Lemma 36.2.7 (Iwasawa decomposition). *The multiplication map gives a homeomorphism*

$$N \times A \times K \xrightarrow{\sim} \mathrm{SL}_2(\mathbb{C}).$$

Proof. We apply the same method as in the proof of Proposition 33.4.2. For surjectivity, we let $z = g(j) = x + yj$, let $n_g = \begin{pmatrix} 1 & -x \\ 0 & 1 \end{pmatrix} \in N$ so that $(n_g g)(j) = yj$; then let $a_g = \begin{pmatrix} 1/\sqrt{y} & 0 \\ 0 & \sqrt{y} \end{pmatrix} \in A$, so $(a_g n_g g)(j) = j$ and $a_g n_g g \in \mathrm{Stab}_{\mathrm{SL}_2(\mathbb{C})}(j) = K$. □

Lemma 36.2.8. *The group* $\mathrm{SL}_2(\mathbb{C})$ *is generated by the subgroups A, N, and the element* $\begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}$, *which acts on* $\mathbf{H}^3$ *by*

$$\begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}(z) = -z^{-1} = \frac{1}{\|z\|^2}(-\overline{x} + yj) \tag{36.2.9}$$

where $\|z\|^2 = \mathrm{nrd}(z) = |x|^2 + y^2$ *for* $z = x + yj$.

Proof. The proof is identical to the one in Lemma 33.4.4. □

Remark 36.2.10. In fact, the generators $\begin{pmatrix} a & 0 \\ 0 & 1/a \end{pmatrix}$ are redundant, but we will not use this fact here.

We are now ready to investigate the consequences of this decomposition for the geometry of hyperbolic space.

Theorem 36.2.11. *The map* (36.2.2) *defines a faithful, transitive action of* $\mathrm{PSL}_2(\mathbb{C})$ *on* $\mathbf{H}^3$ *by isometries.*

Proof. We use the generators in Lemma 36.2.8. The fact that the action is faithful follows directly. For transitivity, we show that $\mathbf{H}^3$ is the orbit of j. If $z = x + yj \in \mathbf{H}^3$ then we first apply a translation $\begin{pmatrix} 1 & -x \\ 0 & 1 \end{pmatrix}$ to reduce to the case $z = yj$ and then reduce to the case of the hyperbolic plane.

Next, we show that $\mathrm{PSL}_2(\mathbb{C}) \hookrightarrow \mathrm{Isom}^+(\mathbf{H}^3)$. Verification that $\mathrm{d}g(s) = \mathrm{d}s$ for g a generator in one of the first two cases of Lemma 36.2.8 is immediate, from the definition of the metric; the third case can be checked directly (Exercise 36.4). Orientation is preserved in each case. $\square$

36.2.12. The group $\mathrm{PSL}_2(\mathbb{C})$ acts transitively on geodesics and consequently on pairs of points at a fixed distance: by the transitive action of $\mathrm{PSL}_2(\mathbb{C})$ on $\mathbf{H}^3$, every point can be mapped to j; and applying an element of $\mathrm{Stab}_{\mathrm{SL}_2(\mathbb{C})}\, j = \mathrm{SU}(2)$, every other point u can be brought to tj with $t \geq 1$, with $\log t = \rho(j, u)$ by the distance in the hyperbolic plane. It follows that

$$\cosh \rho(z, z') = 1 + \frac{|z - z'|^2}{2yy'} = 1 + \frac{|x - x'|^2 + (y - y')^2}{2yy'} \tag{36.2.13}$$

by verifying (36.2.13) in the special case where $z = j$ and $z' = yj$ with $y > 0$, and then using the preceding transitive action and the fact that the right-hand side of (36.2.13) is invariant under the action of $\mathrm{SL}_2(\mathbb{C})$, verified again using the generators in Lemma 36.2.8.

Theorem 36.2.14. *We have*

$$\mathrm{Isom}^+(\mathbf{H}^3) \simeq \mathrm{PSL}_2(\mathbb{C}) \tag{36.2.15}$$

and

$$\mathrm{Isom}(\mathbf{H}^3) \simeq \mathrm{PSL}_2(\mathbb{C}) \rtimes \mathbb{Z}/2\mathbb{Z} \tag{36.2.16}$$

where the nontrivial element of $\mathbb{Z}/2\mathbb{Z}$ *acts by complex conjugation on* $\mathrm{PSL}_2(\mathbb{C})$ *and* $(z, t) \mapsto (\overline{z}, t)$ *on* $\mathbf{H}^3$.

Proof. We argue as in Theorem 33.5.5. Let $\phi \in \mathrm{Isom}(\mathbf{H}^3)$, and let

$$Z = \{yj : y > 0\} \subseteq \mathbf{H}^3.$$

Then Z is a geodesic (see 36.1.5), so $\phi(Z)$ is also a geodesic. By transitivity, there exists an isometry $g \in \mathrm{PSL}_2(\mathbb{C})$ that maps $\phi(j)$ back to j, and we may suppose without loss of generality that $\phi(j) = j$, and arguing as in the case of $\mathbf{H}^2$ we may suppose in fact that ϕ fixes each point of Z. Let $\mathcal{H} = \mathbb{R} + \mathbb{R}j \subseteq \mathbf{H}^3$. Then $\mathcal{H}$ is a geodesic half-plane containing Z, so $\phi(\mathcal{H})$ is as well and must be a vertical half-plane. The

isometric rotation $\begin{pmatrix} e^{i\theta} & 0 \\ 0 & e^{-i\theta} \end{pmatrix}$ fixes Z, and applying such a rotation we may suppose further that ϕ fixes $\mathcal{H}$.

Now let $z = x + yj$ and $\phi(z) = z' = x' + y'j$. Let $r + sj \in \mathcal{H}$. Then

$$\rho(z, r + sj) = \rho(\phi(z), \phi(r + sj)) = \rho(z', r + sj)$$

so from (36.2.13)

$$\frac{|x - r|^2 + (y - s)^2}{2sy} = \frac{|x' - r|^2 + (y' - s)^2}{2sy'};$$

letting $s \to \infty$ we find that $y = y'$ and $|x - r| = |x' - r|$ for all $r \in \mathbb{R}$, thus $\operatorname{Re} x = \operatorname{Re} x'$ and $\operatorname{Im} x = \pm \operatorname{Im} x'$. By continuity, the sign is determined uniquely by g, and we conclude that either $g(z) = z$ or $g(z) = \overline{x} + yj$, as claimed. □

36.2.17. The isometry group $\mathrm{PSL}_2(\mathbb{C})$ also admits a 'purely geometric' definition via the **Poincaré extension**, as follows.

An element $g \in \mathrm{SL}_2(\mathbb{C})$ as a Möbius transformation, induces a biholomorphic map of the Riemann sphere $\mathbb{P}^1(\mathbb{C}) = \mathbb{C} \cup \{\infty\}$. This map can be represented as a composition of an even number (at most four) inversions in circles in $\mathbb{P}^1(\mathbb{C})$, or circles and lines in $\mathbb{C}$ (Exercise 36.2). We have identified $\mathbb{P}^1(\mathbb{C}) = \operatorname{bd} \mathbf{H}^3$ as the boundary, and for each circle in $\mathbb{P}^1(\mathbb{C})$ there is a unique hemisphere in $\mathbf{H}^3$ which intersects $\operatorname{bd} \mathbf{H}^3$ in this circle; if this circle is a line, then we take a vertical half-plane. We then lift the action of $g \in \mathrm{PSL}_2(\mathbb{C})$ one inversion at a time with respect to the corresponding hemisphere or half-plane. It turns out that the action of this product does not depend on the choice of the circles.

To verify that $\mathrm{PSL}_2(\mathbb{C})$ acts by isometries, we need to know that inversion in a hemisphere or vertical half-plane is an isometry of $\mathbf{H}^3$; after observing that the first two types of generators in Lemma 36.2.8 (stretching and translating) are isometries, one reduces to the case of checking that inversion in the unit hemisphere, defined by

$$z \mapsto \frac{z}{\|z\|^2},$$

is an isometry; and this boils down to the same calculation as requested in Exercise 36.4.

36.2.18. We have a similar classification of isometries of $\mathbf{H}^3$ as in the case of $\mathbf{H}^2$ as follows. Let $g \in \mathrm{PSL}_2(\mathbb{C})$.

(i) If $\pm \operatorname{Tr}(g) \in (-2, 2)$, then g is **elliptic**: it has two distinct fixed points in $\operatorname{bd} \mathbf{H}^3$ and fixes every point in the geodesic between them, called its **axis**, acting by (hyperbolic) rotation around its axis.

(ii) If $\pm \operatorname{Tr}(g) \in \mathbb{R} \smallsetminus [-2, 2]$, then g is **hyperbolic**; if $\pm \operatorname{Tr}(g) \in \mathbb{C} \setminus \mathbb{R}$, then g is **loxodromic**. (Some authors combine these two cases.) In these cases, g has two fixed points in $\operatorname{bd} \mathbf{H}^3$ and the line through these two points is stabilized, and g has no fixed point in $\mathbf{H}^3$.

(iii) Finally and otherwise, if $\pm \operatorname{Tr}(g) = \pm 2$, then g is **parabolic**: it has a unique fixed point in $\operatorname{bd} \mathbf{H}^3$ and no fixed point in $\mathbf{H}^3$.

36.3 Unit ball, Lorentz, and symmetric space models

Definition 36.3.1. The **hyperbolic unit ball** is the (open) unit disc

$$\mathbf{D}^3 := \{w = (w_1, w_2, w_3) \in \mathbb{R}^3 : \|w\|^2 < 1\} \tag{36.3.2}$$

equipped with the **hyperbolic metric**

$$\mathrm{d}s := \frac{2\|\mathrm{d}w\|}{1 - \|w\|^2} \tag{36.3.3}$$

and volume

$$\mathrm{d}V := 8\frac{\mathrm{d}w_1\mathrm{d}w_2\mathrm{d}w_3}{(1 - \|w\|)^3}. \tag{36.3.4}$$

The **sphere at infinity** is the boundary

$$\mathrm{bd}\,\mathbf{D}^3 = \{w \in \mathbb{R}^3 : \|w\| = 1\}.$$

36.3.5. The maps

$$\begin{aligned} \phi\colon \mathbf{H}^3 &\xrightarrow{\sim} \mathbf{D}^3 \\ z &\mapsto w = (z - j)(1 - jz)^{-1} \end{aligned} \qquad \begin{aligned} \phi^{-1} : \mathbf{D}^3 &\xrightarrow{\sim} \mathbf{H}^3 \\ w &\mapsto z = (w + j)(1 + jw)^{-1} \end{aligned}$$

define a conformal equivalence between $\mathbf{H}^3$ and $\mathbf{D}^3$ with $j \mapsto \phi(j) = 0$. The hyperbolic metric on $\mathbf{D}^2$ is the pushforward of (induced from) the hyperbolic metric on $\mathbf{H}^3$ via the identification (36.3.5). We find that

$$\cosh \rho(w, w') = 1 + 2\frac{\|w - w'\|^2}{(1 - \|w\|^2)(1 - \|w'\|^2)}. \tag{36.3.6}$$

In the unit ball model, the geodesics are intersections of $\mathbf{D}^3$ of Euclidean circles and straight lines orthogonal to the sphere at infinity, and similarly geodesic planes are intersections of $\mathbf{D}^3$ with Euclidean spheres and Euclidean planes orthogonal to the sphere at infinity.

36.3.7. The isometries of $\mathbf{D}^3$ are obtained by pushforward from $\mathbf{H}^3$. Explicitly, we first identify

$$\begin{aligned} \mathbf{D}^3 &\hookrightarrow \mathbb{H} \\ w &\mapsto w_1 + w_2 i + w_3 j. \end{aligned} \tag{36.3.8}$$

We then define the involution

$$\begin{aligned} {}^* : \mathbb{H} &\to \mathbb{H} \\ \alpha = t + xi + yj + zk &\mapsto k\overline{\alpha}k^{-1} = t + xi + yj - zk \end{aligned} \tag{36.3.9}$$

and the group

$$\mathrm{SU}_2(\mathbb{H}, {}^*) = \left\{ \begin{pmatrix} \alpha & \beta \\ \beta^* & \alpha^* \end{pmatrix} : \alpha, \beta \in \mathbb{H},\ \mathrm{nrd}(\alpha) - \mathrm{nrd}(\beta) = 1 \right\}. \tag{36.3.10}$$

We find that

$$\mathrm{SU}_2(\mathbb{H}, ^*) \simeq \phi\,\mathrm{SL}_2(\mathbb{C})\phi^{-1}$$

with ϕ as in 36.3.5. The group $\mathrm{SU}_2(\mathbb{H}, ^*)$ acts on $\mathbf{D}^3$ by

$$gw = (\alpha w + \beta)(\beta^* w + \alpha^*)^{-1}.$$

36.3.11. Finally, there is the **Lorentz model**

$$\mathbf{L}^3 := \{(t, x) \in \mathbb{R}^4 : -t^2 + x_1^2 + x_2^2 + x_3^2 = -1,\ t > 0\} \tag{36.3.12}$$

with

$$\mathrm{d}s^2 := -\mathrm{d}t^2 + \mathrm{d}x_1^2 + \mathrm{d}x_2^2 + \mathrm{d}x_3^2$$

and orientation-preserving isometries given by the subgroup $\mathrm{SO}^+(3, 1) \leq \mathrm{SO}(3, 1)$ of elements mapping $\mathbf{L}^3$ to itself. The relationship between the Lorentz model and the upper half-space model relies on the exceptional isomorphism of Lie algebras $\mathfrak{so}_{3,1} \simeq \mathfrak{sl}_{2,\mathbb{C}}$ and the double cover $\mathrm{SL}_2(\mathbb{C}) \to \mathrm{SO}^+(3, 1)$.

To conclude, we find the symmetric space model of $\mathbf{H}^3$, analogous to section 34.6.

36.3.13. The group $G := \mathrm{SL}_2(\mathbb{C})$ has the structure of a metric space induced from the usual structure on $\mathrm{M}_2(\mathbb{C}) \simeq \mathbb{C}^4$. Since G acts transitively on $\mathbf{H}^3$, and the stabilizer of j is $K = \mathrm{SU}(2)$,

$$\begin{aligned} G/K = \mathrm{SL}_2(\mathbb{C})/\,\mathrm{SU}(2) &\xrightarrow{\sim} \mathbf{H}^3 \\ gK &\mapsto gj; \end{aligned} \tag{36.3.14}$$

from the Iwasawa decomposition (Lemma 36.2.7), there is a homeomorphism

$$\mathrm{SL}_2(\mathbb{C})/\,\mathrm{SU}(2) \simeq NA.$$

From the identity

$$\|g\|^2 = 2\cosh \rho(j, gj) \tag{36.3.15}$$

for $g \in \mathrm{SL}_2(\mathbb{C})$, proven in the same way as (34.6.5), the map (36.3.14) is a homeomorphism, and even an isometry under the explicit reparametrization (36.3.15) of the metric.

Remark 36.3.16. Similar statements about the unit tangent bundle hold for $\mathrm{PSL}_2(\mathbb{C})$ in place of $\mathrm{PSL}_2(\mathbb{R})$, as in 33.8.2.

Remark 36.3.17. More generally, one defines **hyperbolic upper half-space**

$$\mathbf{H}^n := \{(x, y) \in \mathbb{R}^n \times \mathbb{R} : y > 0\} \text{ with } \mathrm{d}s^2 = \frac{|\mathrm{d}x|^2 + \mathrm{d}y^2}{y^2}.$$

The space $\mathbf{H}^n$ is a uniquely geodesic space and a model for **hyperbolic** n**-space**. The geodesics in $\mathbf{H}^n$ are orthocircles, and via a conformal map. The upper half-space maps isometrically to the (open) unit ball model

$$\mathbf{D}^n := \{x \in \mathbb{R}^n : |x| < 1\} \text{ with } \mathrm{d}s^2 = 4\frac{\mathrm{d}x_1^2 + \cdots + \mathrm{d}x_n^2}{(1 - x_1^2 - \cdots - x_n^2)^2}$$

and the hyperboloid model

$$\mathsf{L}^n := \{(t, x) \in \mathbb{R}^{n+1} : -t^2 + x_1^2 + \cdots + x_n^2 = -1,\ t > 0\}$$

with

$$\mathrm{d}s^2 = -\mathrm{d}t^2 + \mathrm{d}x_1^2 + \cdots + \mathrm{d}x_n^2.$$

These models (and more) are introduced and compared in Cannon–Floyd–Kenyon–Parry [CFKP97], and treated in detail in the works by Benedetti–Petronio [BP92] and Ratcliffe [Rat2006].

Hyperbolic n-space $\mathbf{H}^n$ also admits a symmetric space description, as follows. The group of isometries of $\mathbf{H}^n$ is $\mathrm{SO}(n, 1)$, and the subgroup of orientation-preserving isometries is $\mathrm{SO}^+(n, 1)$, the component of $\mathrm{SO}(n, 1)$ containing the identity matrix. The stabilizer of every point in $\mathbf{H}^n$ is conjugate to $\mathrm{SO}(n)$ (rotation around the origin in the unit ball model, with the fixed point at the origin), and it follows that

$$\mathbf{H}^n \simeq \mathrm{SO}^+(n, 1)/\,\mathrm{SO}(n).$$

36.4 Bianchi groups and Kleinian groups

Theorem 36.4.1. *Let* $G := \mathrm{PSL}_2(\mathbb{C})$ *and let* $\Gamma \leq G$ *be a subgroup. Then the following are equivalent:*

(i) Γ *is discrete (with the subspace topology);*
(ii) *For all* $z \in \mathbf{H}^3$*, we have* $\#\,\mathrm{Stab}_\Gamma(z) < \infty$ *and there exists an open neighborhood* $U \ni z$ *such that* $\gamma U \cap U \neq \emptyset$ *implies* $\gamma \in \mathrm{Stab}_\Gamma(z)$*;*
(iii) *For all compact subsets* $K \subseteq \mathbf{H}^3$*, we have* $K \cap \gamma K \neq \emptyset$ *for only finitely many* $\gamma \in \Gamma$*; and*
(iv) *For all* $z \in \mathbf{H}^3$*, the orbit* $\Gamma z \subseteq \mathbf{H}^3$ *is discrete and* $\#\,\mathrm{Stab}_\Gamma(z) < \infty$.

Moreover, if these equivalent conditions hold, then the quotient $\Gamma\backslash\mathbf{H}^3$ *is Hausdorff, and the quotient map* $\pi \colon \mathbf{H}^3 \to \Gamma\backslash\mathbf{H}^3$ *is a local isometry at all points* $z \in \mathbf{H}^3$ *with* $\mathrm{Stab}_\Gamma(z) = \{1\}$.

Proof. Combine Theorem 34.5.1 and the appropriately modified proof of Proposition 34.7.2. The stabilizer of a point is finite because the stabilizer of $w = 0$ in $\mathrm{SU}_2(\mathbb{H}, {}^*)$ is $\mathrm{SU}(2)$, so its stabilizer in Γ is a discrete subgroup of the compact group $\mathrm{SU}(2)$ thus is necessarily finite (not necessarily cyclic). In particular, a subgroup $\Gamma \leq \mathrm{PSL}_2(\mathbb{C})$ is discrete if and only if the action of Γ on $\mathbf{H}^3$ is wandering, hence proper. □

Definition 36.4.2. A **Kleinian group** is a discrete subgroup of $\mathrm{PSL}_2(\mathbb{C})$.

Let $F \subseteq \mathbb{C}$ be an imaginary quadratic field with ring of integers $R = \mathbb{Z}_F$. Since $R \subseteq \mathbb{C}$ is discrete, and $\mathrm{PSL}_2(R) \subseteq \mathrm{PSL}_2(\mathbb{C})$ is discrete.

Definition 36.4.3. The **Bianchi group** over F is the Kleinian group $\mathrm{PSL}_2(R) \subseteq \mathrm{PSL}_2(\mathbb{C})$.

Remark 36.4.4. The Bianchi groups are so named after work of Bianchi [Bia1892]; he studied them as discrete groups acting on hyperbolic space and found generators in certain cases. For more, see the book by Fine [Fin89].

36.5 Hyperbolic volume

In this section, we consider volumes of hyperbolic polyhedra, following Milnor's chapter in Thurston [Thu97, Chapter 7], published also in Milnor [Milno82, Appendix]; see also the full treatment by Ratcliffe [Rat2006, §10.4].

Definition 36.5.1. The **Lobachevsky function** is defined to be

$$\begin{aligned} \mathcal{L}\colon \mathbb{R} &\to \mathbb{R} \\ \mathcal{L}(\theta) &= -\int_0^\theta \log|2\sin t|\,\mathrm{d}t. \end{aligned} \tag{36.5.2}$$

The Lobachevsky function is also called **Clausen's integral** or more conventionally the **log sine integral**.

36.5.3. The first derivative of the Lobachevsky function is $\mathcal{L}'(\theta) = -\log|2\sin\theta|$ by the fundamental theorem of calculus, so $\mathcal{L}$ attains its maximum value at $\mathcal{L}(\pi/6) = 0.50747\ldots$ and minimum at $\mathcal{L}(5\pi/6) = -\mathcal{L}(\pi/6)$. The second derivative is $\mathcal{L}''(\theta) = -\cot\theta$. A graph of this function is sketched in Figure 36.5.4.

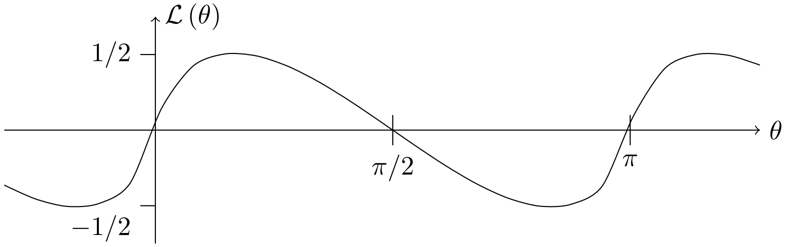


Figure 36.5.4: The Lobachevsky function $\mathcal{L}$

Lemma 36.5.5. *$\mathcal{L}(\theta)$ is odd, periodic with period π, and satisfies the identity*

$$\mathcal{L}(n\theta) = n\sum_{j=0}^{n-1}\mathcal{L}(\theta + j\pi/n) \tag{36.5.6}$$

for all $n \in \mathbb{Z}$.

Proof. Since $\mathcal{L}'(\theta) = -\log|2\sin\theta|$ is an even function and $\mathcal{L}(0) = 0$, we conclude $\mathcal{L}(\theta)$ is an odd function, i.e., $\mathcal{L}(-\theta) = -\mathcal{L}(\theta)$ for all $\theta \in \mathbb{R}$.

Let $n \in \mathbb{Z}$. From

$$z^n - 1 = \prod_{j=0}^{n-1}(z - e^{2\pi i j/n})$$

substituting $z = e^{-2it}$ for $t \in \mathbb{R}$ and using $|e^{2i\theta} - 1| = |1 - e^{2i\theta}| = |2\sin\theta|$ for all $\theta \in \mathbb{R}$ gives

$$|2\sin(nt)| = |1 - e^{2int}| = \prod_{j=0}^{n-1} |e^{-2it} - e^{2\pi ij/n}| = \prod_{j=0}^{n-1} |2\sin(t + j\pi/n)|$$

for all $t \in \mathbb{R}$. Integrating and changing variables $x = nt$ gives

$$\frac{1}{n}\int_0^{\theta} \log|2\sin x|\mathrm{d}x = \sum_{j=0}^{n-1} \int_{j\pi/n}^{\theta + j\pi/n} \log|2\sin x|\mathrm{d}x \tag{36.5.7}$$

which yields

$$\frac{1}{n}\mathcal{L}(n\theta) = \sum_{j=0}^{n-1} \mathcal{L}(\theta + j\pi/n) - \sum_{j=0}^{n-1} \mathcal{L}(j\pi/n) \tag{36.5.8}$$

for all $\theta \in \mathbb{R}$. Plugging in $\theta = \pi/n$ into (36.5.8) yields by telescoping

$$\frac{1}{n}\mathcal{L}(\pi) = \mathcal{L}(\pi) - \mathcal{L}(0) = \mathcal{L}(\pi)$$

so $\mathcal{L}(\pi) = 0$.

Now since $\mathcal{L}'(\theta)$ is periodic with period π and $\mathcal{L}(0) = \mathcal{L}(\pi) = 0$, we conclude that $\mathcal{L}(\theta + \pi) = \mathcal{L}(\theta)$ is also periodic with period π. Finally,

$$\sum_{j=0}^{n-1} \mathcal{L}(j\pi/n) = -\sum_{j=0}^{n-1} \mathcal{L}(-j\pi/n) = -\sum_{j=0}^{n-1} \mathcal{L}((n-j)\pi/n) = \sum_{j=0}^{n-1} \mathcal{L}(j\pi/n)$$

so $\sum_{j=0}^{n-1} \mathcal{L}(j\pi/n) = 0$, and the result follows from (36.5.8). □

Corollary 36.5.9. *We have*

$$\mathcal{L}(2\theta) = 2\mathcal{L}(\theta) + 2\mathcal{L}(\theta + \pi/2) = 2\mathcal{L}(\theta) - 2\mathcal{L}(\pi/2 - \theta).$$

Corollary 36.5.9 is called the **duplication formula** for $\mathcal{L}$.

Proof. Take $n = 2$ in Lemma 36.5.5. □

With the relevant function having been defined, we now return to our geometric application.

Definition 36.5.10. An **ideal tetrahedron** is a tetrahedron whose vertices lie on the sphere at infinity and whose edges are (infinite) geodesics.

36.5.11. An ideal tetrahedron is determined by the three dihedral angles α, β, γ along the edges meeting at any vertex; the sum of these angles is 2π, as the shadow triangle made in $\mathbb{C}$ has angles that sum to 2π. For an illustration, see Figure 36.5.12.

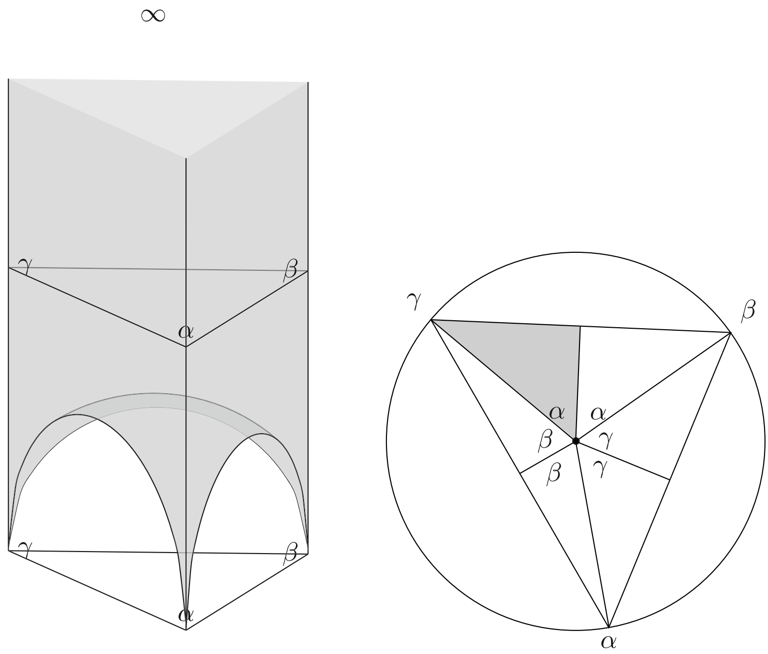


Figure 36.5.12: Ideal tetrahedra and its shadow in $\mathbb{C}$

Proposition 36.5.13. *The volume of an ideal tetrahedron with dihedral angles α, β, γ is $\mathcal{L}(\alpha) + \mathcal{L}(\beta) + \mathcal{L}(\gamma)$.*

Proof. We follow Milnor [Milno82, Appendix, Lemma 2]; see also Thurston(–Milnor) [Thu97, Theorem 7.2.1] and Ratcliffe [Rat2006, Theorem 10.4.10]. We may suppose without loss of generality that one vertex is at ∞ and the finite face lies on the unit sphere. Projecting onto the unit disc in the x-plane, we obtain a triangle inscribed in the unit circle with angles α, β, γ with $\alpha + \beta + \gamma = 2\pi$. We make the simplifying assumption that all three angles are acute (the argument for the case of an obtuse angle is similar). We take the barycentric subdivision of the triangle and add up 6 volumes. We integrate the volume element $\mathrm{d}x_1\mathrm{d}x_2\mathrm{d}y/y^3$ over the region $T(\alpha)$ defined by the inequalities

$$y \geq \sqrt{1 - |x|^2}, \quad 0 \leq x_2 \leq x_1 \tan \alpha, \quad 0 \leq x_1 \leq \cos \alpha. \tag{36.5.14}$$

Integrating with respect to y we have

$$\begin{aligned}\iiint_{T(\alpha)} \frac{\mathrm{d}x_1\mathrm{d}x_2\mathrm{d}y}{y^3} &= \iint -\frac{1}{2}\frac{\mathrm{d}x_1\mathrm{d}x_2}{y^2}\Bigg|_{y=\sqrt{1-|x|^2}}^{\infty} \\ &= -\frac{1}{2}\iint_{\substack{0\leq x_1\leq \cos\alpha \\ 0\leq x_2 \leq x_1 \tan\alpha}} \frac{\mathrm{d}x_1\mathrm{d}x_2}{1 - x_1^2 - x_2^2}\end{aligned}$$

We substitute $x_1 = \cos\theta$, so $\mathrm{d}x_1 = -\sin\theta\mathrm{d}\theta$ and $\pi/2 \geq \theta \geq \alpha$; by partial fractions, we have

$$\int \frac{a\mathrm{d}u}{a^2-u^2} = \frac{1}{2}\log\left|\frac{a+u}{a-u}\right|.$$

So with $a = \sqrt{1-x_1^2} = \sin\theta$, integrating with respect to x_2 gives

$$\begin{aligned}
&\frac{1}{2}\int_\alpha^{\pi/2}\mathrm{d}\theta\int_0^{\cos\theta\tan\alpha}\frac{\sin\theta\mathrm{d}x_2}{\sin^2\theta-x_2^2} \\
&\quad= \frac{1}{4}\int_\alpha^{\pi/2}\mathrm{d}\theta\log\left|\frac{\sin\theta+x_2}{\sin\theta-x_2}\right|_{x_2=0}^{\cos\theta\tan\alpha} \\
&\quad= \frac{1}{4}\int_\alpha^{\pi/2}\log\left|\frac{\sin\theta\cos\alpha+\cos\theta\sin\alpha}{\sin\theta\cos\alpha-\cos\theta\sin\alpha}\right|\mathrm{d}\theta \\
&\quad= \frac{1}{4}\int_\alpha^{\pi/2}\log\left|\frac{2\sin(\theta+\alpha)}{2\sin(\theta-\alpha)}\right|\mathrm{d}\theta \\
&\quad= -\frac{1}{4}\left(\mathcal{L}(\pi/2+\alpha)-\mathcal{L}(2\alpha)-\mathcal{L}(\pi/2-\alpha)\right).
\end{aligned} \tag{36.5.15}$$

Finally, we use the duplication formula (Corollary 36.5.9), which reads

$$\mathcal{L}(2\alpha) = 2\mathcal{L}(\alpha)+\mathcal{L}(\alpha+\pi/2)-\mathcal{L}(\pi/2-\alpha);$$

substituting gives the volume $\mathcal{L}(\alpha)/2$, and summing over the other 5 triangles gives the result. □

We define now a standard tetrahedron for use in computing volumes. Let $T_{\alpha,\gamma}$ be the tetrahedron with one vertex at ∞ and the other vertices A, B, C on the unit hemisphere projecting to A', B', C' in $\mathbb{C}$ with $A' = 0$ to make a Euclidean triangle with angle $\pi/2$ at B' and α at A'. The dihedral angle along the ray from A to ∞ is α. Suppose that the dihedral angle along BC is γ. The acute angles determine the isometry class of $T_{\alpha,\gamma}$, and we call $T_{\alpha,\gamma}$ the **standard tetrahedron** with angles α, γ. For an illustration, see Figure 36.5.18.

Corollary 36.5.16. *We have*

$$\mathrm{vol}(T_{\alpha,\gamma}) = \frac{1}{4}\big(\mathcal{L}(\alpha+\gamma)+\mathcal{L}(\alpha-\gamma)+2\mathcal{L}(\pi/2-\alpha)\big).$$

Proof. One proof realizes the standard tetrahedron as a signed combination of ideal tetrahedra, and uses Proposition 36.5.13. A second proof just repeats the integral (36.5.15) to get

$$\begin{aligned}
\mathrm{vol}(T_{\alpha,\gamma} &= \frac{1}{4}\int_\gamma^{\pi/2}\log\left|\frac{2\sin(\theta+\alpha)}{2\sin(\theta-\alpha)}\right|\mathrm{d}\theta \\
&= -\frac{1}{4}\big(\mathcal{L}(\pi/2+\alpha)-\mathcal{L}(\alpha+\gamma)-\mathcal{L}(\pi/2-\alpha)+\mathcal{L}(\gamma-\alpha)\big)
\end{aligned} \tag{36.5.17}$$

which rearranges to give the result. □

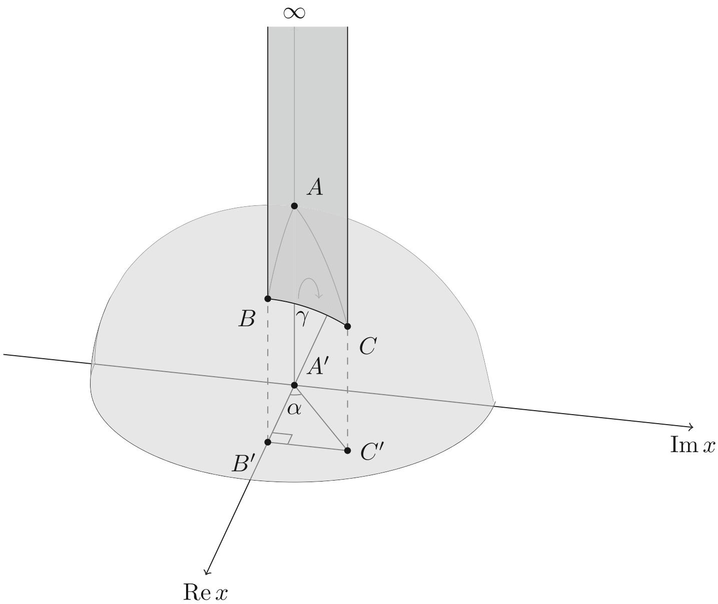


Figure 36.5.18: Standard tetrahedron

36.5.19. By Exercise 36.11, we have the Fourier expansion

$$\mathcal{L}(\theta) = \frac{1}{2}\sum_{n=1}^{\infty} \frac{\sin(2n\theta)}{n^2}. \tag{36.5.20}$$

The series (36.5.20) converges rather slowly, but twice integrating the Laurent series expansion for $\cot\theta$ as in Exercise 36.12 gives

$$\mathcal{L}(\theta) = \theta\left(1 - \log|2\theta| + \sum_{n=1}^{\infty} \frac{|B_{2n}|}{4n}\frac{(2\theta)^{2n+1}}{(2n+1)!}\right)$$

where

$$\frac{x}{e^x - 1} = \sum_{k=0}^{\infty} B_k \frac{x^k}{k!} = 1 - \frac{x}{2} + \frac{1}{6}\frac{x^2}{2!} - \frac{1}{30}\frac{x^4}{4!} + \dots$$

so $|B_2| = 1/6$, $|B_4| = 1/30$, etc. are the **Bernoulli numbers**.

36.6 Picard modular group

In this section, analogous to the case of the classical modular group $\mathrm{PSL}_2(\mathbb{Z})$ we consider the special case of a full Bianchi group with $K = \mathbb{Q}(i)$.

Definition 36.6.1. The group $\mathrm{PSL}_2(\mathbb{Z}[i])$ is called the **(full) Picard modular group**.

Throughout this section, we write $\Gamma = \mathrm{PSL}_2(\mathbb{Z}[i])$. In order to understand the structure of the group Γ, we follow the same script as in section 35.1, and first we seek a fundamental set.

Proposition 36.6.2. *Let*

$$⌑ = \{z = x + yj \in \mathbf{H}^3 : |z|^2 \geq 1,\ |\mathrm{Re}\, x| \leq 1/2,\ 0 \leq \mathrm{Im}\, x \leq 1/2\}. \tag{36.6.3}$$

Then ⌑ *is a fundamental set for* $\Gamma \circlearrowright \boldsymbol{H}^3$, *and* $\mathrm{PSL}_2(\mathbb{Z}[i])$ *is generated by the elements*

$$\begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}, \begin{pmatrix} 1 & i \\ 0 & 1 \end{pmatrix}, \begin{pmatrix} i & 0 \\ 0 & -i \end{pmatrix}, \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}. \tag{36.6.4}$$

The set ⌑ in Proposition 36.6.2 is displayed in Figure 36.6.5.

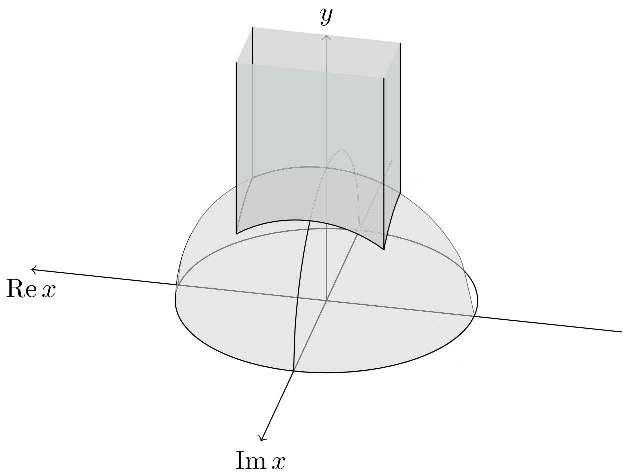


Figure 36.6.5: A fundamental set ⌑ for the Picard group $\mathrm{PSL}_2(\mathbb{Z}[i]) \circlearrowright \mathbf{H}^3$

Proof. First, we show that for all $z \in \mathbf{H}^3$, there exists a word γ in the matrices (36.6.4) such that $\gamma z \in$ ⌑ via an explicit reduction algorithm. Let $z = x + yj \in \mathbf{H}^3$. Recalling the action (36.2.6), the element $\begin{pmatrix} 1 & b \\ 0 & 1 \end{pmatrix}$ for $b \in \mathbb{Z}[i]$ act by translation $z \mapsto z + b$, so repeatedly applying matrices from the first two among (36.6.4), we may suppose that $|\mathrm{Re}\, x|, |\mathrm{Im}\, x| \leq 1/2$. Then applying the element $\begin{pmatrix} i & 0 \\ 0 & -i \end{pmatrix}$, which acts by $z \mapsto (iz)(-i)^{-1} = i^2(x - yj) = -x + yj$, we may suppose $\mathrm{Im}\, x \geq 0$. Now if $z \in$ ⌑, which is to say $\|z\|^2 \geq 1$, we are done. Otherwise, $\|z\|^2 < 1$, and we apply the matrix $\gamma = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}$, which by (36.2.9) acts by

$$\gamma z = z' = \frac{1}{\|z\|^2}(-\overline{x} + yj) = x' + y'j$$

and $y' = y/\|z\|^2 > y$, and we repeat. Since Γz is discrete, this terminates after finitely many steps: the set

$$\Gamma z \cap \left\{ z' = x' + y'j \in \mathbf{H}^3 : |\operatorname{Re} x'|, |\operatorname{Im} x'| \le 1/2,\ y \le y' \le 1 \right\}$$

is discrete and compact, hence finite.

Next, if $z, z' \in$ ⌑ and $z \in$ int(⌑) with $z' = \gamma z$ for $\gamma \in \Gamma$, then $\gamma = 1$ and $z = z'$; this can be proven directly as in Lemma 35.1.10 (the details are requested in Exercise 36.7). It follows that the matrices (36.6.4) generate $\mathrm{PSL}_2(\mathbb{Z}[i])$ as in Lemma 35.1.12, taking instead $z_0 = 2j \in$ int(⌑). ☐

A slightly more convenient set of generators, together with the gluing relations they provide on the fundamental set, is given in Figure 36.6.7.

Remark 36.6.6. By a deeper investigation into the structure of the fundamental set ⌑, in chapter 37 we will find a presentation for Γ as

$$\Gamma \simeq \langle \gamma_1, \gamma_2, \gamma_3, \gamma_4 \mid \gamma_1^2 = \gamma_2^2 = \gamma_4^2 = 1,$$
$$(\gamma_3\gamma_1)^3 = (\gamma_3\gamma_2)^2 = (\gamma_3\gamma_4)^2 = (\gamma_2\gamma_1)^2 = (\gamma_4\gamma_1)^3 = 1\rangle.$$

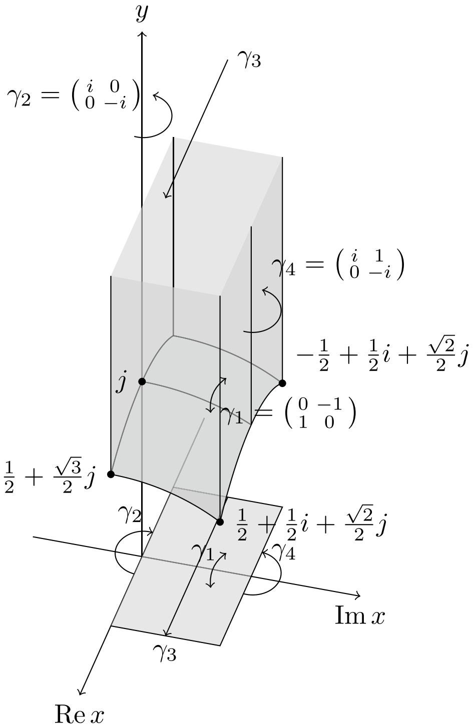


Figure 36.6.7: Generators for Picard group $\mathrm{PSL}_2(\mathbb{Z}[i]) \circlearrowright \mathbf{H}^3$

36.6.8. We compute the volume of this fundamental domain using formulas from the previous section. First, we use symmetry to triangulate (tetrahedralize) ⌑, as in Figure 36.6.9.

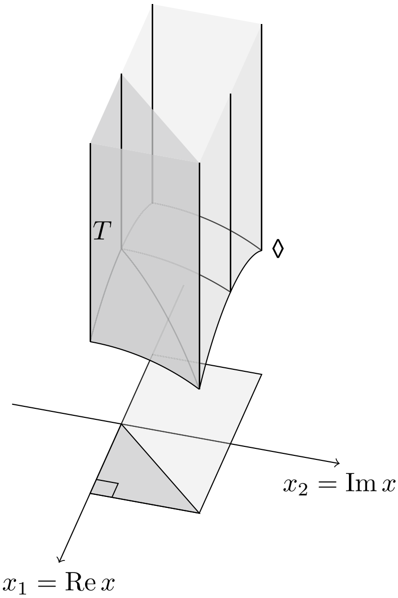


Figure 36.6.9: Triangulation of ⌑

Let

$$T := \{z = x_1 + x_2 i + yj \in \mathbf{H}^3 : 0 \le x_2 \le x_1 \le 1/2,\ x_1^2 + x_2^2 + y^2 \ge 1\}. \tag{36.6.10}$$

Applying the symmetries of Γ, we see that $\operatorname{vol}(⌑) = 4 \operatorname{vol}(T)$. We have $T = T_{\alpha,\gamma}$ a standard tetrahedron, with $\alpha = \pi/4$ and dihedral angle $\gamma = \pi/3$.

Now by the hard-earned volume formula (Corollary 36.5.16) we have

$$\operatorname{vol}(T) = \frac{1}{4}\big(\mathcal{L}(\pi/4 + \pi/3) + \mathcal{L}(\pi/4 - \pi/3) + 2\mathcal{L}(\pi/4)\big). \tag{36.6.11}$$

By Lemma 36.5.5 with $n = 3$, we have

$$\frac{1}{3}\mathcal{L}(3\pi/4) = \mathcal{L}(\pi/4) + \mathcal{L}(\pi/4 + \pi/3) + \mathcal{L}(\pi/4 + 2\pi/3); \tag{36.6.12}$$

since

$$\begin{gathered} \mathcal{L}(3\pi/4) = \mathcal{L}(\pi - \pi/4) = -\mathcal{L}(\pi/4) \\ \mathcal{L}(\pi/4 + 2\pi/3) = \mathcal{L}(\pi/4 + 2\pi/3 - \pi) = \mathcal{L}(\pi/4 - \pi/3), \end{gathered}$$

substituting (36.6.12) into (36.6.11) gives

$$\operatorname{vol}(T) = \frac{1}{4}\left(2 - 1 - \frac{1}{3}\right)\mathcal{L}(\pi/4) = \frac{1}{6}\mathcal{L}(\pi/4) = 0.07633\ldots \tag{36.6.13}$$

and
$$\operatorname{vol}(\text{⌑}) = 4\operatorname{vol}(T) = \frac{2}{3}\mathcal{L}(\pi/4) = 0.30532\ldots$$

36.6.14. We conclude with a beautiful consequence of this volume calculation, giving a preview of the volume formula we will prove later. By the Fourier expansion (36.5.20), we have
$$\mathcal{L}(\pi/4) = \frac{1}{2}\sum_{n=1}^{\infty}\frac{\sin(n\pi/2)}{n^2} = \frac{1}{2}\sum_{n=1}^{\infty}\frac{\chi(n)}{n^2} \tag{36.6.15}$$
where
$$\chi(n) = \begin{cases} 0, & \text{if } 2 \mid n; \\ 1, & \text{if } n \equiv 1 \pmod 4; \\ -1, & \text{if } n \equiv -1 \pmod 4). \end{cases}$$
is the nontrivial *Dirichlet character* modulo 4. We can analytically continue the sum (36.6.15) to $\mathbb{C}$ via the L-series
$$L(s,\chi) = \sum_{n=1}^{\infty}\frac{\chi(n)}{n^s}$$
for $s \in \mathbb{C}$ with $\operatorname{Re} s > 1$ whose general study was the heart of Part III of this text. Here, we can just observe that $L(2,\chi) = 2\mathcal{L}(\pi/4) = 0.915965\ldots$, so
$$\operatorname{vol}(\text{⌑}) = \operatorname{vol}(\Gamma\backslash\mathbf{H}^3) = \frac{1}{3}L(2,\chi) = 0.30532\ldots$$
More generally, the volume of the quotient by a Bianchi group is connected to an L-value attached to the associated imaginary field; we will pursue this topic further in chapter 39.

Exercises

1. For $z \in \mathbf{H}^3$ and $g = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in \mathrm{PSL}_2(\mathbb{C})$, show that
$$(az+b)(cz+d)^{-1} = (zc+d)^{-1}(za+b).$$

▹ 2. **Inversion** in the circle of radius r in $\mathbb{C}$ centered at the origin is defined by the map
$$z \mapsto r^2\left(\frac{z}{|z|^2}\right) = \frac{r^2}{\overline{z}},$$
as in Figure 36.6.16.

Sending $0 \mapsto \infty$ and $\infty \mapsto 0$ under this map, we obtain an anti-holomorphic map $\mathbb{P}^1(\mathbb{C}) \to \mathbb{P}^1(\mathbb{C})$. Inversion in a line in $\mathbb{C}$ is reflection in the line.
Verify that every element of $\mathrm{PSL}_2(\mathbb{C})$ can be written as a composition of at most four inversions in circles and lines in $\mathbb{C}$ (or equivalently, by stereographic projection, circles in $\mathbb{P}^1(\mathbb{C})$).

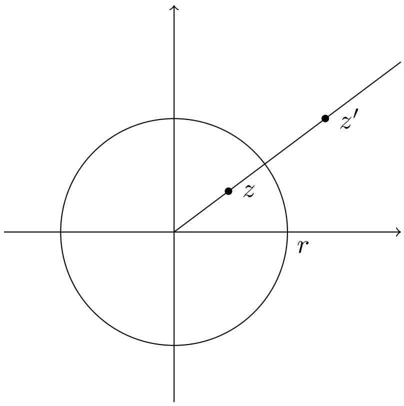


Figure 36.6.16: Inversion in the circle of radius r

▹ 3. Verify (36.2.3) for the action of $\mathrm{SL}_2(\mathbb{C})$ on $\mathbf{H}^3 \subseteq \mathbb{H}$.

▹ 4. Let $g = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix} \in \mathrm{SL}_2(\mathbb{C})$. Show that g acts on $\mathbf{H}^3$ by

$$gz = \frac{1}{\|z\|^2}(-\overline{x} + yj)$$

where $\|z\|^2 = |x|^2 + y^2$, and that g is a hyperbolic isometry.

5. Let $\Gamma \leq \mathrm{PSL}_2(\mathbb{C})$ be a subgroup (with the subspace topology). Show that Γ has a wandering action on $\mathbf{H}^3$ if and only if Γ is discrete (cf. Proposition 34.7.2).

6. Show that the reduction algorithm in Proposition 36.6.2 recovers the Euclidean algorithm for $\mathbb{Z}[i]$ in a manner analogous to Exercise 35.4 for $\mathbb{Z}$.

▹ 7. Consider the fundamental set ⌑ for $\Gamma = \mathrm{PSL}_2(\mathbb{Z}[i])$ (36.6.3). Show that if $z, z' \in$ ⌑ and $z \in \mathrm{int}($⌑$)$ with $z' = \gamma z$ for $\gamma \in \Gamma$, then $\gamma = 1$ and $z = z'$ (cf. Lemma 35.1.10).

8. Let $\omega = e^{2\pi i/3} \in \mathbb{C}$. The field $\mathbb{Q}(\omega) = \mathbb{Q}(\sqrt{-3})$ is Euclidean under the norm, just like $\mathbb{Q}(i)$. Give a description of a fundamental domain for the group $\mathrm{PSL}_2(\mathbb{Z}[\omega])$ analogous to Proposition 36.6.2. [The fields $\mathbb{Q}(\sqrt{d})$ with $d < 0$ are Euclidean if and only if $d = -3, -4, -7, -8, -11$, so similar—but increasingly difficult—arguments can be given in each of these cases. See Fine [Fin89, §4.3] for presentations.]

▹ 9. We consider hyperplane bisectors in $\mathbf{H}^3$ (cf. Exercise 33.8). Let $z_1, z_2 \in \mathbf{H}^3$ be distinct. Let

$$H(z_1, z_2) = \{z \in \mathbf{H}^3 : \rho(z, z_1) \leq \rho(z, z_2)\}$$

be the locus of points as close to z_1 as to z_2, and let

$$L(z_1, z_2) = \mathrm{bd}\, H(z_1, z_2).$$

Show that $H(z_1, z_2)$ is a convex half-space (for every two points in the half-space, the geodesic between them is contained in the half-space), and that

$$L(z_1, z_2) = \{z \in \mathbf{H}^3 : \rho(z, z_1) = \rho(z, z_2)\}$$

is geodesic and equal to the perpendicular bisector of the geodesic segment from z_1 to z_2.

10. Prove the duplication formula for the Lobachevsky function $\mathcal{L}(\theta)$ using the double angle formula, given that $\mathcal{L}(\pi/2) = 0$.

▹ 11. In this exercise, we prove the Fourier expansion

$$\mathcal{L}(\theta) = \frac{1}{2}\sum_{n=1}^{\infty}\frac{\sin(2n\theta)}{n^2}. \tag{36.6.17}$$

(a) Define the **dilogarithm function** by

$$\mathrm{Li}_2(z) = \sum_{n=1}^{\infty}\frac{z^n}{n^2};$$

show that this series converges for $|z| < 1$ and that

$$\mathrm{Li}_2(z) = -\int_0^z \frac{\log(1-w)}{w}\mathrm{d}w.$$

(b) Prove that

$$2i\mathcal{L}(\theta) = \mathrm{Li}_2(e^{2i\theta}) - \frac{\pi^2}{6} + \pi\theta - \theta^2. \tag{36.6.18}$$

[Hint: Differentiate both sides for $0 < \theta < \pi$*, using the limiting value as* $\theta \to 0$ *to compute the limiting value* $\mathrm{Li}_2(1) = \pi^2/6$*.]*

(c) Take imaginary parts of (36.6.18) to prove (36.6.17).

▹ 12. As in 36.5.19, we define the $B_k \in \mathbb{Q}$ for $k \geq 0$ by

$$\frac{x}{e^x - 1} = \sum_{k=0}^{\infty} B_k \frac{x^k}{k!} = 1 - \frac{x}{2} + \frac{1}{6}\frac{x^2}{2!} - \frac{1}{30}\frac{x^4}{4!} + \dots. \tag{36.6.19}$$

(a) Plug in $x = 2iz$ into (36.6.19) to get

$$z \cot z = 1 + \sum_{k=2}^{\infty} B_k \frac{(2iz)^k}{k!}.$$

(b) Integrate twice in (a) to prove

$$\mathcal{L}(\theta) = \theta\left(1 - \log|2\theta| + \sum_{n=1}^{\infty}\frac{|B_{2n}|}{4n}\frac{(2\theta)^{2n+1}}{(2n+1)!}\right).$$

[See also Exercise 40.3.]

This chapter is licensed under the terms of the Creative Commons Attribution-NonCommercial 4.0 International License (http://creativecommons.org/licenses/by-nc/4.0/), which permits any noncommercial use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license and indicate if changes were made.

The images or other third party material in this chapter are included in the chapter's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the chapter's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.

Chapter 37

Fundamental domains

We have seen in sections 35.1 and 36.6 that understanding a nice fundamental set for the action of a discrete group Γ is not only useful to visualize the action of the group by selecting representatives of the orbits, but it is also instrumental for many other purposes, including understanding the structure of the group itself. In this chapter, we pursue a general construction of nice fundamental domains for the action of a discrete group of isometries.

37.1 ▹ Dirichlet domains for Fuchsian groups

In this introductory section, we preview the results in this chapter specialized to the case of Fuchsian groups. Let $\Gamma \subset \mathrm{PSL}_2(\mathbb{R})$ be a Fuchsian group; then Γ is discrete, acting properly by isometries on the hyperbolic plane $\mathbf{H}^2$, with metric $\rho(\cdot,\cdot)$ and hyperbolic area μ.

A natural way to produce fundamental sets that provide appealing tessellations of $\mathbf{H}^2$ is to select in each orbit the points closest to a fixed point $z_0 \in \mathbf{H}^2$, as follows.

Definition 37.1.1. The **Dirichlet domain** for Γ centered at $z_0 \in \mathbf{H}^2$ is

$$\mathcal{D}(\Gamma; z_0) = \{z \in \mathbf{H}^2 : \rho(z, z_0) \le \rho(\gamma z, z_0) \text{ for all } \gamma \in \Gamma\}.$$

As the group Γ willst vary, we suppress the dependence on Γ and often write simply $\mathcal{D}(z_0) = \mathcal{D}(\Gamma; z_0)$.

37.1.2. The set $\mathcal{D}(z_0)$ is an intersection

$$\mathcal{D}(z_0) = \bigcap_{\gamma \in \Gamma} H(\gamma; z_0) \qquad (37.1.3)$$

where

$$H(\gamma; z_0) := \{z \in \mathbf{H}^2 : \rho(z, z_0) \le \rho(\gamma z, z_0) = \rho(z, \gamma^{-1} z_0)\}. \qquad (37.1.4)$$

In particular, since each $H(\gamma; z_0)$ is closed, we conclude from (37.1.3) that $\mathcal{D}(z_0)$ is closed.

© The Author(s) 2021

J. Voight, *Quaternion Algebras*, Graduate Texts in Mathematics 288,
https://doi.org/10.1007/978-3-030-56694-4_37

The sets $H(\gamma; z_0)$ can be further described as follows. If $z_0 = \gamma^{-1} z_0$, then $H(\gamma; z_0) = \mathbf{H}^2$. So suppose $z_0 \neq \gamma^{-1} z_0$. Then by Exercise 33.8, $H(\gamma; z_0)$ is a (half!) half-plane consisting of the set of points as close to z_0 as $\gamma^{-1} z_0$, and $H(\gamma; z_0)$ is convex: if two points lie in the half-plane then so does the geodesic segment between them. The boundary

$$\operatorname{bd} H(\gamma; z_0) = L(\gamma; z_0) := \{z \in \mathbf{H}^2 : \rho(z, z_0) = \rho(z, \gamma^{-1} z_0)\}$$

is the geodesic perpendicular bisector of the geodesic segment from z_0 to $\gamma^{-1} z_0$, as in Figure 37.1.5.

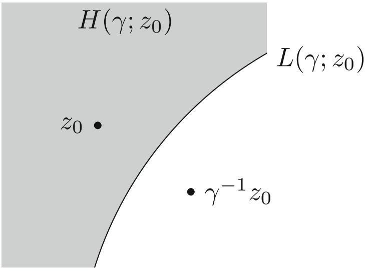


Figure 37.1.5: The half-plane $H(\gamma; z_0)$ and its boundary $L(\gamma; z_0)$

From the description in 37.1.2, the sketch of a Dirichlet domain looks like Figure 37.1.6.

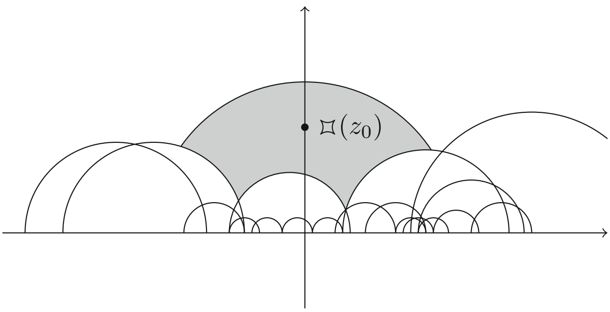


Figure 37.1.6: Sketch of a Dirichlet domain

The fundamental sets we have seen are in fact examples of Dirichlet domains.

Example 37.1.7. We claim that the Dirichlet domain for $\Gamma = \mathrm{PSL}_2(\mathbb{Z})$ centered at $z_0 = 2i$ is in fact the fundamental set for Γ introduced in section 35.1, i.e.,

$$⌑(\Gamma; z_0) = ⌑(2i) = \{z \in \mathbf{H}^2 : |\operatorname{Re} z| \leq 1/2,\ |z| \geq 1\}. \tag{37.1.8}$$

Recall the generators $S, T \in \Gamma$ with $Sz = -1/z$ and $Tz = z + 1$. By (33.5.3)

$$\cosh \rho(z, 2i) = 1 + \frac{|z - 2i|^2}{4 \operatorname{Im} z}.$$

Let $z \in \mathbf{H}^2$. Visibly

$$\rho(z, 2i) \le \rho(Tz, 2i) \quad \Leftrightarrow \quad \operatorname{Re} z \ge -1/2 \tag{37.1.9}$$

or put another way

$$H(T; 2i) = \{z \in \mathbf{H}^2 : \operatorname{Re} z \ge -1/2\}.$$

Similarly, $H(T^{-1}; 2i) = \{z \in \mathbf{H}^2 : \operatorname{Re} z \le 1/2\}$. Equivalently, the geodesic perpendicular bisector between $2i$ and $2i \pm 1$ are the lines $\operatorname{Re} z = \pm 1/2$.

In the same manner, we find that

$$\begin{aligned} \rho(z, 2i) \le \rho(Sz, 2i) \quad &\Leftrightarrow \quad \frac{|z - 2i|^2}{\operatorname{Im} z} \le \frac{|(-1/z) - 2i|^2}{\operatorname{Im}(-1/z)} = \frac{4|z|^2|z - i/2|^2}{|z|^2 \operatorname{Im} z} \\ &\Leftrightarrow \quad |z - 2i|^2 \le 4|z - i/2|^2 \end{aligned}$$

so

$$\rho(z, 2i) \le \rho(Sz, 2i) \quad \Leftrightarrow \quad |z| \ge 1 \tag{37.1.10}$$

and $H(S; 2i) = \{z \in \mathbf{H}^2 : |z| \ge 1\}$. To see this geometrically, the geodesic between $2i$ and $S(2i) = (1/2)i$ is along the imaginary axis with midpoint at i, and so the perpendicular bisector $L(S; 2i)$ is the unit semicircle.

The containment ($\subseteq$) in (37.1.8) then follows directly from (37.1.9)–(37.1.10). Conversely, we show the containment ($\supseteq$) for the interior—since $\text{⌑}(2i)$ is closed, this implies the full containment. Let $z \in \mathbf{H}^2$ have $|\operatorname{Re} z| < 1/2$ and $|z| > 1$, and suppose that $z \notin \text{⌑}(2i)$; then there exists $\gamma \in \mathrm{PSL}_2(\mathbb{Z})$ such that $z' = \gamma z$ has $\rho(z', 2i) < \rho(z, 2i)$, without loss of generality (replacing z' by Sz' or Tz') we may suppose $|\operatorname{Re} z'| \le 1/2$ and $|z'| \ge 1$; but then by Lemma 35.1.10, we conclude that $z' = z$, a contradiction.

(Note that the same argument works with $z_0 = ti$ for all $t \in \mathbb{R}_{>1}$.)

With this example in hand, we see that Dirichlet domains have quite nice structure. To make this more precise, we upgrade our notion of fundamental set (Definition 34.1.14) as follows.

Definition 37.1.11. A fundamental set ⌑ for Γ is **locally finite** if for each compact set $K \subset \mathbf{H}^2$, we have $\gamma K \cap \text{⌑} \neq \emptyset$ for only finitely many $\gamma \in \Gamma$.

A **fundamental domain** for $\Gamma \circlearrowright \mathbf{H}^2$ is a fundamental set $\text{⌑} \subseteq \mathbf{H}^2$ such that $\mu(\operatorname{bd} \text{⌑}) = 0$.

The first main result of this section is as follows (Theorem 37.5.3).

Theorem 37.1.12. *Let* $z_0 \in \mathbf{H}^2$ *satisfy* $\operatorname{Stab}_\Gamma(z_0) = \{1\}$. *Then the Dirichlet domain* $\text{⌑}(\Gamma; z_0)$ *is a connected, convex, locally finite fundamental domain for* Γ *with geodesic boundary.*

By *geodesic boundary* we mean that the boundary $\operatorname{bd} \text{⌑}(z_0)$ is a finite or countable union of geodesic segments. As for the hypothesis: for a compact set $K \subset \mathbf{H}^2$, there are only finitely many points $z \in K$ such that $\operatorname{Stab}_\Gamma(z) \neq \{1\}$, indeed there are only finitely many $\gamma \in \Gamma$ such that $\gamma K \cap K \neq \emptyset$ (as Γ is discrete), and every such $\gamma \neq 1$ has at most one fixed point in $\mathbf{H}^2$ (Lemma 33.4.6).

37.2 ▹ Ford domains

In this section, we reinterpret Dirichlet domains in the unit disc $\mathbf{D}^2$, as it is more convenient to compute and visualize distances this model. Let $z_0 \in \mathbf{H}^2$. We apply the map (33.7.3)

$$\phi\colon \mathbf{H}^2 \to \mathbf{D}^2$$
$$w = \frac{z - z_0}{z - \overline{z_0}}$$

with $z_0 \mapsto \phi(z_0) = w_0 = 0$. Then by (33.7.6),

$$\rho(w, 0) = \log \frac{1 + |w|}{1 - |w|} = 2 \tanh^{-1} |w| \tag{37.2.1}$$

is an increasing function of $|w|$.

Example 37.2.2. The Dirichlet domain from Example 37.1.7 looks like Figure 37.2.3, drawn in $\mathbf{D}^2$ (with $z_0 = 2i$).

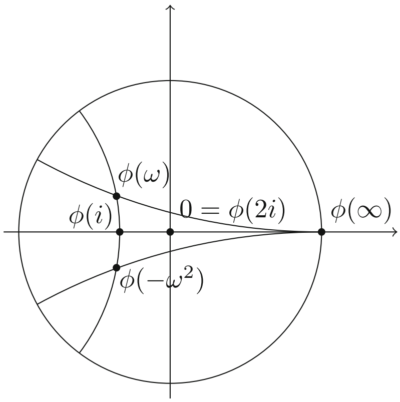


Figure 37.2.3: Dirichlet domain in $\mathbf{D}^2$

Let $\Gamma \leq \mathrm{PSL}_2(\mathbb{R})$ be a Fuchsian group, and (recalling 33.7.8) to ease notation, we identify Γ with Γ^ϕ. We analogously define a Dirichlet domain $\text{⌑}(w_0)$ for a Fuchsian group Γ centered at $w_0 \in \mathbf{D}^2$ and

$$\phi(\text{⌑}(z_0)) = \text{⌑}(w_0) \subset \mathbf{D}^2$$

if $\phi(z_0) = w_0$. In particular, the statement of Theorem 37.1.12 applies equally well to $\text{⌑}(w_0) \subseteq \mathbf{D}^2$.

For simplicity (and without loss of generality), we only consider the case where $w_0 = 0$, and then from (37.2.1)

$$\text{⌑}(\Gamma; 0) = \{w \in \mathbf{D}^2 : |w| \leq |\gamma w| \text{ for all } \gamma \in \Gamma\}. \tag{37.2.4}$$

Let

$$g = \begin{pmatrix} \overline{d} & \overline{c} \\ c & d \end{pmatrix} \in \mathrm{PSU}(1,1) \circlearrowright \mathbf{D}^2$$

with $c, d \in \mathbb{C}$ satisfying $|d|^2 - |c|^2 = 1$.

37.2.5. We now pursue a tidy description of the set (37.2.4). From (37.2.1), we have $\rho(w,0) \leq \rho(gw,0)$ if and only if

$$|w| \leq \left|\frac{\overline{d}w + \overline{c}}{cw + d}\right|; \tag{37.2.6}$$

expanding out (37.2.6) and with a bit of patience (Exercise 37.5), we see that this is equivalent to simply

$$|cw + d| \geq 1.$$

But we can derive this more conceptually, as follows. The hyperbolic metric (Definition 33.7.1) on $\mathbf{D}^2$ is invariant, so

$$\mathrm{d}s = \frac{|\mathrm{d}w|}{(1-|w|)^2} = \frac{|\mathrm{d}(gw)|}{(1-|gw|)^2} = \mathrm{d}(gs);$$

so, by the chain rule,

$$\left|\frac{\mathrm{d}g}{\mathrm{d}w}(w)\right| = \left(\frac{1-|gw|}{1-|w|}\right)^2.$$

Therefore

$$|w| \leq |gw| \quad \Leftrightarrow \quad \left|\frac{\mathrm{d}g}{\mathrm{d}w}(w)\right| = \frac{1}{|cw+d|^2} \leq 1 \quad \Leftrightarrow \quad |cw+d| \geq 1. \tag{37.2.7}$$

The equivalence (37.2.7) also gives $\rho(w,0) = \rho(gw,0)$ if and only if $\left|\frac{\mathrm{d}g}{\mathrm{d}w}(w)\right| = 1$, i.e., g acts as a *Euclidean* isometry at the point w (preserving the length of tangent vectors in the Euclidean metric). So we are led to make the following definition.

Definition 37.2.8. The **isometric circle** of g is

$$I(g) = \left\{w \in \mathbb{C} : \left|\frac{\mathrm{d}g}{\mathrm{d}w}(w)\right| = 1\right\} = \{w \in \mathbb{C} : |cw+d| = 1\}.$$

We have $c = 0$ if and only if $g(0) = 0$ if and only if $g \in \mathrm{Stab}_{\mathrm{PSU}(1,1)}(0)$, and in this case, $gw = (\overline{d}/d)w$ with $|\overline{d}/d| = 1$ is rotation about the origin. Otherwise, $c \neq 0$, and then $I(g)$ is a circle with radius $1/|c|$ and center $-d/c \in \mathbb{C}$.

37.2.9. Summarizing, for all $g \in \mathrm{PSU}(1,1)$, we have

$$\rho(w,0) \begin{Bmatrix} < \\ = \\ > \end{Bmatrix} \rho(gw,0) \ \text{ according as } w \in \begin{cases} \mathrm{ext}(I(g)), \\ I(g), \\ \mathrm{int}(I(g)). \end{cases}$$

In particular,

$$\maltese(\Gamma;0) = \bigcap_{\gamma \in \Gamma - \mathrm{Stab}_\Gamma(0)} \mathrm{cl}\,\mathrm{ext}\, I(\gamma).$$

This characterization is due to Ford [For72, Theorem 7, §20].

This description of a Dirichlet domain as the intersection of the exteriors of isometric circles is due to Ford, and so we call a Dirichlet domain in $\mathbf{D}^2$ centered at 0 a **Ford domain**, as in Figure 37.2.10.

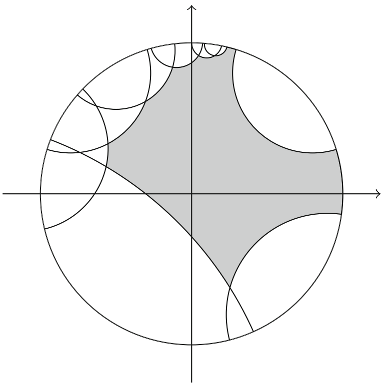


Figure 37.2.10: A typical Ford domain

Remark 37.2.11. In the identification $\mathbf{H}^2 \to \mathbf{D}^2$, the preimage of isometric circles corresponds to the corresponding perpendicular bisector; this is the simplification provided by working in $\mathbf{D}^2$ (the map ϕ is a hyperbolic isometry, whereas isometric circles are defined by a Euclidean condition).

37.3 ▹ Generators and relations

Continuing with our third and final survey section focused on Fuchsian groups, we consider applications to the structure of a Fuchsian group Γ. For more, see Beardon [Bea95, §9.3] and Katok [Kat92, §3.5].

Let ⌑ $=$ ⌑$(\Gamma; z_0)$ be a Dirichlet domain centered at $z_0 \in \mathbf{H}^2$. A consequence of the local finiteness of a Dirichlet domain is the following theorem (Theorem 37.4.2).

Theorem 37.3.1. Γ *is generated by the set*

$$\{\gamma \in \Gamma : ⌑ \cap \gamma ⌑ \neq \emptyset\}.$$

So by Theorem 37.3.1, to find generators, we must look for "overlaps" in the tessellation provided by ⌑. If $z \in ⌑ \cap \gamma ⌑$ with $\gamma \in \Gamma \smallsetminus \{1\}$, then there exists $z, z' \in ⌑$ such that $z = \gamma z'$, and hence

$$\rho(z, z_0) \leq \rho(\gamma^{-1} z, z_0) = \rho(z', z_0) \leq \rho(\gamma z', z_0) = \rho(z, z_0) \qquad (37.3.2)$$

(twice applying the defining property of ⌑), so equality holds and (viz. 37.1.2) $z \in \operatorname{bd} ⌑$. Since the boundary of ⌑ is geodesic, to understand generators we should organize the matching provided along the geodesic boundary of ⌑.

We will continue to pass between $\mathbf{H}^2$ and $\mathbf{D}^2$, as convenient.

37.3.3. A **side** is a geodesic segment of positive length of the form $⌑ \cap \gamma⌑$ with $\gamma \in \Gamma \smallsetminus \{1\}$. If two sides intersect in D^2, the point of intersection is called a **vertex** of $⌑$; equivalently, a vertex is a single point of the form $⌑ \cap \gamma⌑ \cap \gamma'⌑$ with $\gamma, \gamma' \in \Gamma$.

However, we make the following convention (to simplify the arguments below): if $L = ⌑ \cap \gamma⌑$ is a maximal geodesic subset of $⌑$ and $\gamma^2 = 1$, or equivalently if $\gamma L = L$, then γ fixes the midpoint of L, and we consider L to be the union of *two* sides that meet at the vertex equal to the midpoint. The representation of a side as $⌑ \cap \gamma⌑$ is unique when $\gamma^2 \neq 1$.

Because $⌑$ is locally finite, there are only finitely many vertices in a compact neighborhood (Exercise 37.7).

An **ideal vertex** is a point of the closure of $⌑$ in D^{2*} that is the intersection of the closure of two sides in D^{2*}, as in Figure 37.3.4.

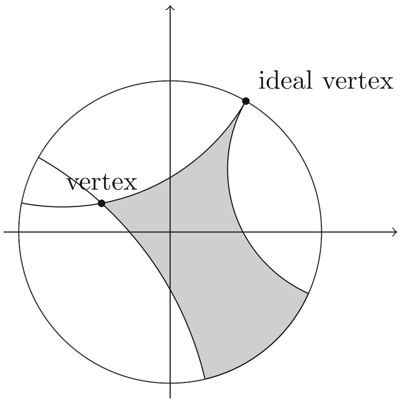


Figure 37.3.4: A vertex and an ideal vertex

37.3.5. Let S denote the set of sides of $⌑$. We define a labeled equivalence relation on S by

$$P = \{(\gamma, L, L^*) : L^* = \gamma(L)\} \subset \Gamma \times (S \times S). \tag{37.3.6}$$

We say that P is a **side pairing** if P induces a partition of S into pairs, and we denote by $G(P)$ the projection of P to Γ. Since $(\gamma, L, L^*) \in P$ implies $(\gamma^{-1}, L^*, L) \in P$, we conclude that $G(P)$ is closed under inverses.

Lemma 37.3.7. *A Dirichlet domain* $⌑$ *has a side pairing* P.

Proof. Let L be a side with $L \subseteq ⌑ \cap \gamma⌑$ for a unique γ. Recalling the convention in 37.3.3, if $\gamma^2 \neq 1$, then equality holds, and

$$\gamma^{-1}L = ⌑ \cap \gamma^{-1}⌑ = L^* \neq L$$

so by uniqueness, the equivalence class of L contains only L, L^*. If $\gamma^2 = 1$, then L meets $\gamma L = L^*$ at the fixed point of γ, and again the equivalence class of L contains on L, L^*. In either case, we conclude that P (37.3.6) is a side pairing. □

37.3.8. We now provide a **standard picture** of $⌑$ in a neighborhood of a point $w \in \operatorname{bd} ⌑$.

Because ⌑ is locally finite, there is an an open neighborhood of w and finitely many distinct $\delta_0, \delta_1, \dots, \delta_n \in \Gamma$ with $\delta_0 = 1$ such that $U \subseteq \bigcup_{i=0}^{n} \delta_i$⌑ and $w \in \delta_i$⌑ for all i. Shrinking U if necessary, we may suppose that U contains no vertices of ⌑ except possibly for w and intersects no sides of δ_i⌑ except those that contain w. Therefore, we have a picture as in Figure 37.3.9.

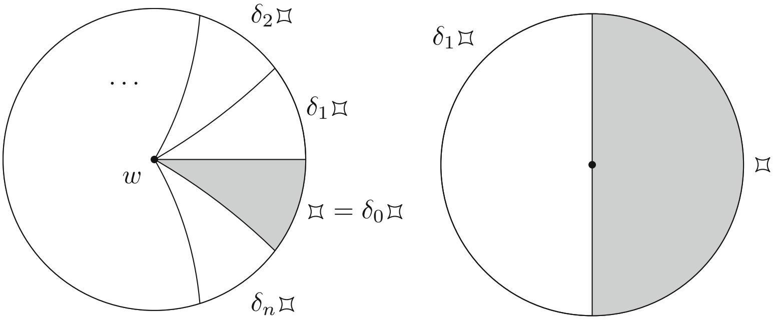


Figure 37.3.9: Standard picture

When $n = 1$, then either w can be either a vertex (fixed point of δ_1) or not.

37.3.10. Let $v = v_1$ be a vertex of ⌑. The standard picture in a neighborhood of v can be reinterpreted as in Figure 37.3.11.

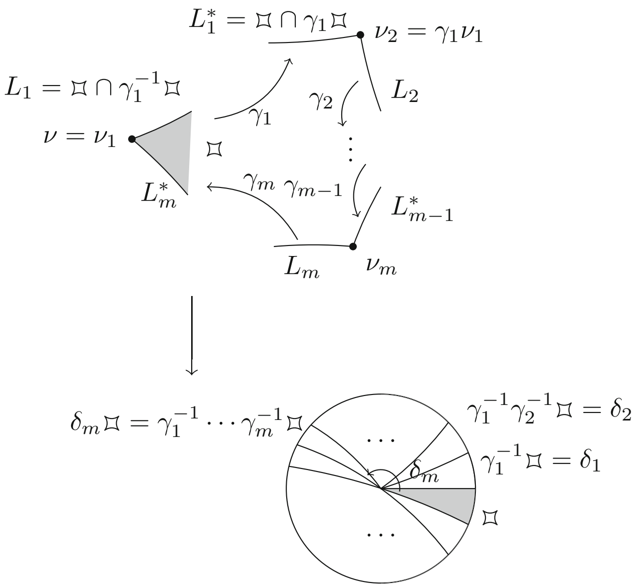


Figure 37.3.11: The standard picture in a neighborhood of v

Let L_1 be the side containing v_1 traveling clockwise from the interior. Then by the side pairing (Lemma 37.3.7), there is a paired side L_1^* with $L_1^* = \gamma_1 L_1$ and $\gamma_1 \in G(P)$. (In fact, then $L_1 = ⌑ \cap \gamma_1^{-1}⌑$ and $L_1^* = ⌑ \cap \gamma_1 ⌑$.) Let $v_2 = \gamma_1 v_1$. Then v_2 is a vertex of ⌑, and so is contained in a second side L_2. Continuing in this way, with $v_{i+1} = \gamma_i v_i$, by local finiteness we find after finitely many steps a final side L_m^* with next vertex $v_m = v_1$.

In terms of the standard picture (Figure 37.3.9), we see that $\delta_1 = \gamma_1^{-1}$ and by induction $\delta_i = (\gamma_i \cdots \gamma_1)^{-1}$, since $(\gamma_i \cdots \gamma_1)(v_1) = v_{i+1}$. Thus $\gamma_{i+1} = \delta_{i+1}^{-1}\delta_i$ for $i = 0, \dots, m-1$. Let $\delta = \delta_m$. Then $\delta(v) = v$, and δ acts by counterclockwise hyperbolic rotation with fixed point v—and m is the smallest nonzero index with this property. It follows that for all $0 \leq j \leq n$, writing $j = qm + i$ with $q \geq 0$ and $0 \leq i < m$ we have $\delta_j = \delta_m^q \delta_i$, and in particular that $m \mid (n+1)$. Similarly, $⌑ \cap \Gamma v_1 = \{v_1, \dots, v_{m-1}\}$.

Let $e = (n+1)/m$. Then $\delta^e = 1$, and we call this relation the **vertex cycle relation** for v. If $v' = \gamma v$, then the vertex cycle relation for v' is the conjugate relation $(\gamma^{-1}\delta\gamma)^e = \gamma^{-1}\delta^e\gamma = 1$. Let $R(P)$ be the set of (conjugacy classes of) cycle relations arising from Γ-orbits of vertices in ⌑.

Example 37.3.12. We compute the set $R(P)$ of cycle relations for $\Gamma = \mathrm{PSL}_2(\mathbb{Z})$. The two Γ-orbits of vertices for ⌑ are represented by i and ρ. The vertex i is a fixed point of $\delta_1 = S$, and we obtain the vertex cycle relation $S^2 = 1$, and $S = S^{-1}$.

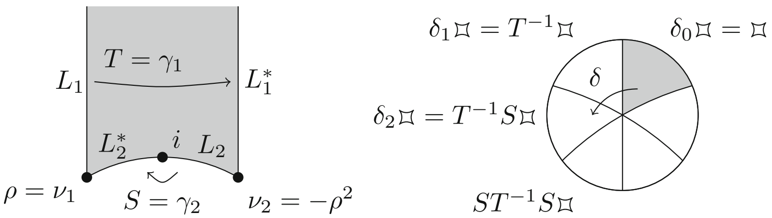


Figure 37.3.13: The cycle relation at ρ

At the vertex ρ, we have a picture as in Figure 37.3.13.
We find that $\delta_1 = T^{-1}$ and $\delta = \delta_2 = T^{-1}S$, and $e = 6/2 = 3$ so $\delta^3 = (T^{-1}S)^3 = 1$. Taking inverses (so δ^{-1} acting instead clockwise), we find the relation $(ST)^3 = 1$.

Proposition 37.3.14. *The set $G(P)$ generates Γ with $R(P)$ a set of defining relations. In other words, the free group on $G(P)$ modulo the normal subgroup generated by the relations $R(P)$ is isomorphic to Γ via the natural evaluation map.*

Proof. Let $\Gamma^* \leq \Gamma$ be the subgroup generated by $G(P)$. By Theorem 37.3.1, we need to show that if $⌑ \cap \gamma⌑ \neq \emptyset$ then $\gamma \in \Gamma^*$. So let $w \in ⌑ \cap \gamma⌑$ with $\gamma \in \Gamma - \{1\}$. We refer to the standard picture (see 37.3.8); we have $\gamma = \delta_j$ for some j. For all $i = 0, 1, \dots, n$, the intersection $\delta_i⌑ \cap \delta_{i+1}⌑$ is a side of $\delta_{i+1}⌑$, so $⌑ \cap \delta_{i+1}^{-1}\delta_i⌑$ is a side of ⌑, and thus $\delta_{i+1}^{-1}\delta_i \in G(P)$ is a side pairing element. Since $\delta_0 = 1$, by induction we find that $\delta_i \in \Gamma^*$ for all i, so $\gamma = \delta_j \in \Gamma^*$ as claimed.

We now turn to relations, and we give an algorithmic method for rewriting a relation in terms of the cycle relations. Let $\gamma_1\gamma_2 \cdots \gamma_k = 1$ be a relation with each $\gamma_i \in G(P)$,

and let $z_i = (\gamma_1 \cdots \gamma_i)z_0$ for $i = 1, \dots, k$. Exactly because $\gamma_1 \in G(P)$, we have that ⌑ and γ_1⌑ share a side, and since ⌑ is connected, we can draw a path $z_0 \to z_1$ through the corresponding side. Continuing in this way, we end up with a path $z_0 \to z_k = z_0$, hence a closed loop, as in Figure 37.3.15.

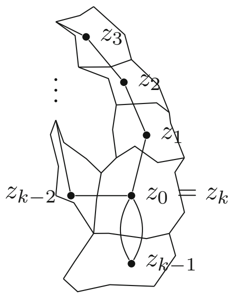


Figure 37.3.15: A closed loop

Let V be the intersection of the Γ orbit of the vertices of ⌑ with the interior of the loop; by local finiteness, this intersection is a finite set, and we proceed by induction on its cardinality. The proof boils down to the fact that this loop retracts onto the loops around vertices obtained from cycle relations, as $\mathbf{H}^2$ is simply connected.

If the path from $z_0 \to z_1$ crosses the same side as the path $z_{k-1} \to z_k = z_0$, then $z_1 = z_{k-1}$ and so $\gamma_k^{-1} = \gamma_1$, since $\mathrm{Stab}_\Gamma(z_0) = \{1\}$ (see Figure 37.3.16).

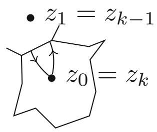


Figure 37.3.16: Simplifying a relation: setup

Conjugating the relation by γ_1 and repeating if necessary, we may suppose that $\gamma_k^{-1} \neq \gamma_1$, so $z_{k-1} \neq z_1$; the set V is conjugated so it remains the same size. In particular, if V is empty, then this shows that the original relation is conjugate to the trivial relation.

Otherwise, the path $z_0 \to z_1$ crosses a side and there is a unique vertex v on this side that is interior to the loop (working counterclockwise): see Figure 37.3.17.

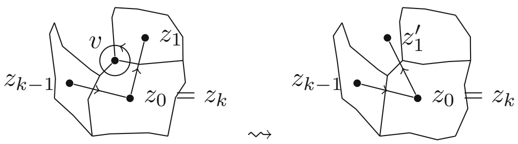


Figure 37.3.17: Simplifying a relation: using the cycle relation

The cycle relation $\delta^e = 1$ at v traces a loop around v, and without loss of generality we may suppose $\delta^e = \alpha\gamma_1$ with α a word in $G(P)$. Therefore, substituting this relation into the starting relation, we obtain a relation $\gamma_2 \cdots \gamma_k(\delta^e\alpha)$ with one fewer interior vertex; the result then follows by induction. $\square$

In section 37.5, we consider a partial converse to Proposition 37.3.14, due to Poincaré: given a convex hyperbolic polygon with a side pairing that satisfies certain conditions, there exists a Fuchsian group Γ with the given polygon as a fundamental domain.

37.4 Dirichlet domains

We now consider the construction of Dirichlet fundamental domains in a general context. Let (X, ρ) be a complete, locally compact geodesic space. In particular, X is connected, and by the theorem of Hopf–Rinow (see Theorem 33.2.9), closed balls in X are compact.

Let Γ be a discrete group of isometries acting properly on X. Right from the get go, we prove our first important result: we exhibit generators for a group based on a fundamental set with a basic finiteness property.

Definition 37.4.1. Let $A \subseteq X$. We say A is **locally finite** for Γ if for each compact set $K \subset X$, we have $\gamma K \cap A \neq \emptyset$ for only finitely many $\gamma \in \Gamma$.

The value of a locally finite fundamental set is explained by the following theorem.

Theorem 37.4.2. *Let* ⌑ *be a locally finite fundamental set for* Γ. *Then* Γ *is generated by the set*

$$\{\gamma \in \Gamma : ⌑ \cap \gamma ⌑ \neq \emptyset\}. \tag{37.4.3}$$

Proof. Let $\Gamma^* \leq \Gamma$ be the subgroup of Γ generated by the elements (37.4.3). We want to show $\Gamma^* = \Gamma$.

For all $x \in X$, by Theorem 37.4.18, there exists $\gamma \in \Gamma$ such that $\gamma x \in ⌑$. If there is another $\gamma' \in \Gamma$ with $\gamma' x \in ⌑$, then

$$\gamma' x \in ⌑ \cap \gamma'\gamma^{-1} ⌑$$

so $\gamma'\gamma^{-1} \in \Gamma^*$ and in particular $\Gamma^*\gamma = \Gamma^*\gamma'$. In this way, we define a map

$$\begin{aligned} f\colon X &\to \Gamma^*\backslash\Gamma \\ x &\mapsto \Gamma^*\gamma \end{aligned} \tag{37.4.4}$$

for all $\gamma \in \Gamma$ such that $\gamma x \in ⌑$.

We now show that f is locally constant. Let $x \in X$. Since ⌑ is locally finite, for every compact neighborhood $K \ni x$ we can write $K \subseteq \bigcup_i \gamma_i ⌑$ with a finite union, and by shrinking K we may suppose that $x \in \gamma_i ⌑$ for all i. In particular, $f(x) = \Gamma^*\gamma_i^{-1}$ for any i. But then if $y \in K$, then $y \in \gamma_i ⌑$ for some i, so $f(y) = \Gamma^*\gamma_i^{-1} = f(x)$. Thus f is locally constant.

But X is connected so every locally constant function is in fact constant and therefore f takes only the value Γ^*. Let $\gamma \in \Gamma$ and let $x \in ⌑$. Then

$$\Gamma^* = f(x) = f(\gamma^{-1}x) = \Gamma^*\gamma$$

so $\gamma \in \Gamma^*$, and the proof is complete. $\square$

We now seek a locally finite fundamental set with other nice properties: we will choose in each Γ-orbit the closest points to a fixed point $x_0 \in X$. So we first must understand the basic local properties of intersections of these half-spaces (as in 37.1.2).

37.4.5. For $x_1, x_2 \in X$, define the closed **Leibniz half-space**

$$H(x_1, x_2) = \{x \in X : \rho(x, x_1) \leq \rho(x, x_2)\}. \tag{37.4.6}$$

If $x_1 = x_2$, then $H(x_1, x_2) = X$. If $x_1 \neq x_2$, then $H(x_1, x_2)$ consists of the set of points as close to x_1 as x_2, so

$$\operatorname{int} H(x_1, x_2) = \{x \in X : \rho(x, x_1) < \rho(x, x_2)\}. \tag{37.4.7}$$

and

$$\operatorname{bd} H(x_1, x_2) = L(x_1, x_2) = \{x \in X : \rho(x, x_1) = \rho(x, x_2)\}$$

is called the **hyperplane bisector** (or **equidistant hyperplane** or **separator**) between x_1 and x_2.

Remark 37.4.8. In this generality, unfortunately hyperplane bisectors are not necessarily geodesic (Exercise 37.10).

Definition 37.4.9. A set $A \subseteq X$ is **star-shaped** with respect to $x_0 \in A$ if for all $x \in A$, the geodesic between x and x_0 belongs to A.

A set $A \subseteq X$ that is star-shaped is path connected, so connected.

Lemma 37.4.10. *A Leibniz half-plane $H(x_1, x_2)$ is star-shaped with respect to x_1.*

Proof. Let $x \in H(x_1, x_2)$ and let y be a point along the unique geodesic from x to x_1. Then

$$\rho(x_1, y) + \rho(y, x) = \rho(x_1, x).$$

If $y \notin H(x_1, x_2)$, then $\rho(x_2, y) < \rho(x_1, y)$, and so

$$\rho(x_2, x) \leq \rho(x_2, y) + \rho(y, x) < \rho(x_1, y) + \rho(y, x) = \rho(x_1, x)$$

contradicting that $x \in H(x_1, x_2)$. So $y \in H(x_1, x_2)$ as desired. $\square$

Now let $x_0 \in X$.

Definition 37.4.11. The **Dirichlet domain** for Γ centered at $x_0 \in X$ is

$$⌑(\Gamma; x_0) = \{z \in X : \rho(x, x_0) \leq \rho(\gamma x, x_0) \text{ for all } \gamma \in \Gamma\}.$$

We often abbreviate $⌑(x_0) = ⌑(\Gamma; x_0)$.

37.4.12. Since $\rho(\gamma x, x_0) = \rho(x, \gamma^{-1} x_0)$,

$$⌑(x_0) = \bigcap_{\gamma \in \Gamma} H(x_0, \gamma^{-1} x_0);$$

each half-space is closed and star-shaped with respect to x_0, so the same is true of $⌑(x_0)$. In particular, $⌑(x_0)$ is connected.

A Dirichlet domain satisfies a basic finiteness property, as follows.

Lemma 37.4.13. *If $A \subset X$ is a bounded set, then $A \subseteq H(\gamma; x_0)$ for all but finitely many $\gamma \in \Gamma$.*

In particular, for all $x \in X$ we have $x \in H(\gamma; x_0)$ for all but finitely many $\gamma \in \Gamma$.

Proof. Since A is bounded,

$$\sup(\{\rho(x, x_0) : x \in A\}) = r < \infty.$$

By Theorem 34.5.1, the orbit Γx_0 is discrete and $\#\operatorname{Stab}_\Gamma(x_0) < \infty$; since closed balls are compact by assumption, there are only finitely many $\gamma \in \Gamma$ such that

$$\rho(\gamma x_0, x_0) = \rho(x_0, \gamma^{-1} x_0) \leq 2r$$

and for all remaining $\gamma \in \Gamma$ and all $x \in A$,

$$\rho(x, \gamma^{-1} x_0) \geq \rho(x_0, \gamma^{-1} x_0) - \rho(x, x_0) > 2r - r = r \geq \rho(x, x_0)$$

so $x \in H(\gamma; x_0)$, as in Figure 37.4.14.

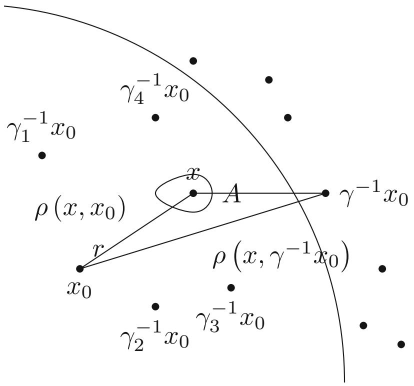


Figure 37.4.14: The bounded set A and the orbit of x_0

This conclude the proof. □

37.4.15. Arguing in a similar way as in Lemma 37.4.13, one can show (Exercise 37.8): if K is a compact set, then $K \cap L(\gamma; x_0) \neq \emptyset$ for only finitely many $\gamma \in \Gamma$.

Lemma 37.4.16. *We have*

$$\operatorname{int} ⌑(x_0) = \{x \in ⌑ : \rho(x, x_0) < \rho(\gamma x, x_0) \text{ for all } \gamma \in \Gamma \smallsetminus \operatorname{Stab}_\Gamma(x_0)\}.$$

It follows from Lemma 37.4.16 that $x \in \operatorname{bd} ⌑(x_0)$ if and only if there exists $\gamma \in \Gamma \smallsetminus \operatorname{Stab}_\Gamma(x_0)$ such that $\rho(x, x_0) = \rho(\gamma x, x_0)$.

Proof. Let $x \in ⌑$, and let $U \ni x$ be a bounded open neighborhood of x. By Lemma 37.4.13, we have $U \subseteq H(\gamma; x_0)$ for all but finitely many $\gamma \in \Gamma$, so

$$U \cap ⌑ = U \cap \bigcap_i H(x_0, \gamma_i^{-1} x_0)$$

the intersection over finitely many $\gamma_i \in \Gamma$ with $\gamma_i \notin \mathrm{Stab}_\Gamma(x_0)$ (see Figure 37.4.17).

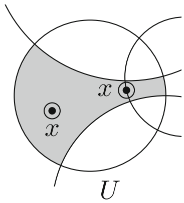


Figure 37.4.17: The neighborhood U

Thus

$$U \cap \mathrm{int}(⌑) = U \cap \bigcap_i \mathrm{int}\, H(x_0, \gamma_i^{-1} x_0).$$

The lemma then follows from (37.4.7). □

The first main result of this chapter is the following theorem.

Theorem 37.4.18. *Let $x_0 \in X$, and suppose $\mathrm{Stab}_\Gamma(x_0) = \{1\}$. Then $⌑(\Gamma; x_0)$ is a locally finite fundamental set for Γ that is star-shaped with respect to x_0 and whose boundary consists of hyperplane bisectors.*

Specifically, in a bounded set A, by Lemma 37.4.13

$$A \cap \mathrm{bd}\, ⌑(\Gamma; x_0) \subseteq \bigcup_i L(x_0, \gamma_i x_0)$$

for finitely many $\gamma_i \in \Gamma \smallsetminus \{1\}$.

Proof. Abbreviate $⌑ = ⌑(x_0)$. We saw that $⌑$ is (closed and) star-shaped with respect to x_0 in 37.4.12.

Now we show that $X = \bigcup_\gamma \gamma ⌑$. Let $x \in X$. The orbit Γx is discrete, so the distance

$$\rho(\Gamma x, x_0) = \inf(\{\rho(\gamma x, x_0) : \gamma \in \Gamma\}) \tag{37.4.19}$$

is minimized at some point $\gamma x \in ⌑$ with $\gamma \in \Gamma$. Thus $⌑(x_0)$ contains at least one point from every Γ-orbit, and consequently .

We now refer to Lemma 37.4.16. Since X is complete, this lemma implies that $\mathrm{cl}(\mathrm{int}(⌑)) = ⌑$. And $\mathrm{int}(⌑) \cap \mathrm{int}(\gamma ⌑) = \emptyset$ for all $\gamma \in \Gamma \smallsetminus \{1\}$, because if $x, \gamma x \in \mathrm{int}(⌑)$ with $\gamma \neq 1$ then

$$\rho(x, x_0) < \rho(\gamma x, x_0) < \rho(\gamma^{-1}(\gamma x), x_0) = \rho(x, x_0), \tag{37.4.20}$$

a contradiction.

Finally, we show that X is locally finite. It suffices to check this for a closed disc $K \subseteq X$ with center x_0 and radius $r \in \mathbb{R}_{\geq 0}$. Suppose that γK meets ⌑ with $\gamma \in \Gamma$; then by definition there is $x \in$ ⌑ such that $\rho(x_0, \gamma^{-1}x) \leq r$. Then

$$\rho(x_0, \gamma^{-1}x_0) \leq \rho(x_0, \gamma^{-1}x) + \rho(\gamma^{-1}x, \gamma^{-1}x_0) \leq r + \rho(x, x_0).$$

Since $x \in$ ⌑, we have $\rho(x, x_0) \leq \rho(\gamma^{-1}x, x_0) \leq r$, so

$$\rho(x_0, \gamma^{-1}x_0) \leq r + r = 2r.$$

This setup can be seen in Figure 37.4.21.

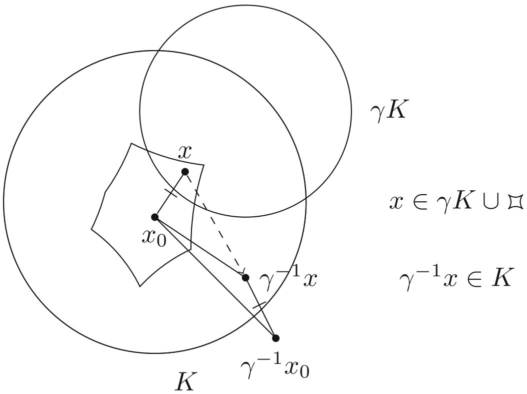


Figure 37.4.21: ⌑ is locally finite and star-shaped

For the same reason as in Lemma 37.4.13, this can only happen for finitely many $\gamma \in \Gamma$. □

Remark 37.4.22. Dirichlet domains are sometimes also called *Voronoi domains*.

37.5 Hyperbolic Dirichlet domains

We now specialize to the case $X = \mathcal{H}$ where $\mathcal{H} = \mathbf{H}^2$ or $\mathcal{H} = \mathbf{H}^3$ with volume μ; then Γ is a Fuchsian or Kleinian group, respectively.

Definition 37.5.1. A **fundamental domain** for $\Gamma \circlearrowright X$ is a connected fundamental set ⌑ $\subseteq X$ such that $\mu(\mathrm{bd}$ ⌑$) = 0$.

(A *domain* in topology is sometimes taken to be an open connected set; one also sees *closed domains*, and our fundamental domains are taken to be of this kind.)

We now turn to Dirichlet domains in this context.

37.5.2. In Theorem 37.4.18, the hypothesis that $\mathrm{Stab}_\Gamma(z_0) = \{1\}$ is a very mild hypothesis. If K is a compact set, then since ⌑ is locally finite, the set of points $z \in K$ with $\mathrm{Stab}_\Gamma(z) \neq \{1\}$ is a finite set of points when $\mathcal{H} = \mathbf{H}^2$ and a finite set of points together with finitely many geodesic axes when $\mathcal{H} = \mathbf{H}^3$.

In spite of 37.5.2, we prove a slightly stronger and more useful version of Theorem 37.4.18, as follows. If $\Gamma_0 = \operatorname{Stab}_\Gamma(z_0)$ is nontrivial, we modify the Dirichlet domain by intersecting with a fundamental set for Γ_0; the simplest way to do this is just to choose another point which is not fixed by an element of Γ_0 and intersect.

Theorem 37.5.3. *Let $z_0 \in \mathcal{H}$, let $\Gamma_0 = \operatorname{Stab}_\Gamma(z_0)$, and let $u_0 \in \mathcal{H}$ be such that* $\operatorname{Stab}_{\Gamma_0}(u_0) = \{1\}$. *Then*

$$D(\Gamma; z_0) \cap D(\Gamma_0; u_0)$$

is a connected, convex, locally finite fundamental domain for Γ with geodesic boundary in $\mathcal{H}$.

Proof. Abbreviate

$$D = D(\Gamma; z_0) \cap D(\Gamma_0; u_0).$$

First, we show that $z_0 \in D$: we have $z_0 \in D(\Gamma; z_0)$, and by Theorem 37.4.18, $D(\Gamma_0; u_0)$ is a fundamental set for Γ_0 so there exists $\gamma_0 \in \Gamma_0$ such that $\gamma_0 u_0 = u_0 \in D(\Gamma_0; u_0)$.

Now we show that D is a fundamental set for Γ. First we show $\mathcal{H} = \bigcup_\gamma \gamma D$. Let $z \in \mathcal{H}$, and let $\gamma \in \Gamma$ be such that $\rho(\gamma z, z_0)$ is minimal as in (37.4.19). Let $\gamma_0 \in \Gamma_0$ such that $\gamma_0(\gamma z) \in D(\Gamma_0; u_0)$. Then

$$\rho(\gamma_0(\gamma z), z_0) = \rho(\gamma z, z_0)$$

so $\gamma_0\gamma z \in D$. And $\operatorname{int}(D) \cap \operatorname{int}(\gamma D) = \emptyset$ for all $\gamma \in \Gamma \smallsetminus \{1\}$, because if $z, \gamma z \in \operatorname{int}(D)$ with $\gamma \neq 1$, then either $\gamma \notin \Gamma_0$ in which case we obtain a contradiction as in (37.4.20), or $\gamma \in \Gamma_0 - \{1\}$ and then we arrive at a contradiction from the fact that $D(\Gamma_0; u_0)$ is a fundamental set.

We conclude by proving the remaining topological properties of D. We know that D is locally finite, since it is the intersection of two locally finite sets. We saw in 37.1.2 that the Leibniz half-spaces in $\mathbf{H}^2$ are convex with geodesic boundary, and the same is true in $\mathbf{H}^3$ by Exercise 36.9. It follows that D is convex, as the intersection of convex sets. Thus

$$\operatorname{bd} D \subseteq \bigcup_{\gamma \in \Gamma \smallsetminus \{1\}} L(z_0, \gamma^{-1} z_0)$$

is geodesic and measure zero, since $L(z_0, \gamma^{-1} z_0)$ intersects a compact set for only finitely many γ by 37.4.15. □

37.6 Poincaré's polyhedron theorem

Continuing with the notation from the previous section, we now turn to a partial converse for Theorem 37.5.3 for $\mathcal{H}$; we need one additional condition. Let D be a convex (finite-sided) hyperbolic polyhedron equipped with a side pairing P.

37.6.1. First suppose $D \subseteq \mathbf{H}^2$. For a vertex v of D, let $\vartheta(D, v)$ be the interior angle of D at v. We say that D **satisfies the cycle condition** if for all vertices v of D there exists $e \in \mathbb{Z}_{>0}$ such that

$$\sum_{v_i \in \Gamma v \cap D} \vartheta(D, v_i) = \frac{2\pi}{e}.$$

Put another way, $⌑$ satisfies the cycle condition if the sum of the interior angles for a Γ-orbit of vertices as in the standard picture is an integer submultiple of 2π.

Now suppose $⌑ \subseteq \mathbf{H}^3$. Now we work with edges: for an edge ℓ of $⌑$, let $\vartheta(⌑, \ell)$ be the interior angle of $⌑$ at ℓ. We say that $⌑$ **satisfies the cycle condition** if for all edges ℓ of $⌑$ there exists $e \in \mathbb{Z}_{>0}$ such that

$$\sum_{\ell_i \in \Gamma e \cap ⌑} \vartheta(⌑, \ell_i) = \frac{2\pi}{e}.$$

Lemma 37.6.2. *Let $⌑$ be a Dirichlet domain. Then $⌑$ satisfies the cycle condition.*

Proof. We explain the case where $⌑ \subseteq \mathbf{H}^2$; the case of $\mathbf{H}^3$ is similar. Let v be a vertex of $⌑$. Referring to the standard picture 37.3.8,

$$2\pi = \sum_{j=0}^{n} \vartheta(\delta_j ⌑, v) = \sum_{j=0}^{n} \vartheta(⌑, \delta_j^{-1} v).$$

In 37.3.10, we proved that δ_m acts by hyperbolic rotation around v and has order $e = (n+1)/m$, and

$$\sum_{j=0}^{n} \vartheta(⌑, \delta_j^{-1} v) = e \sum_{i=0}^{m-1} \vartheta(⌑, \delta_i^{-1} v) = e \sum_{i=0}^{m-1} \vartheta(⌑, v_i)$$

with $⌑ \cap \Gamma v = \{v_1, \dots, v_{m-1}\}$. Combining these two equations, we see that the cycle condition is satisfied. □

37.6.3. There is another condition at certain points at infinity that a fundamental domain must satisfy. We say a point $z \in \operatorname{bd} \mathcal{H}$ is a **infinite vertex** if z lies in the intersection of two faces and is tangent to both. We define a sequence of tangency vertices analogous to the cycle transformations to get an **infinite vertex sequence** and a **infinite vertex transformation**. We say that the side/face pairing is **complete** if the tangency vertex transformation is parabolic.

Theorem 37.6.4 (Poincaré polygon theorem). *Let $⌑$ be a convex finite-sided hyperbolic polygon/polyhedron with a side/face-pairing P. Suppose that $⌑$ satisfies the cycle condition and P is complete.*

Then the group $\Gamma := \langle G(P) \rangle$ generated by side/face-pairing elements is a Fuchsian/Kleinian group, $⌑$ is a fundamental domain for Γ, and $R(P)$ forms a complete set of relations for $G(P)$.

Proof. Unfortunately, it is beyond the scope of this book to give a complete proof of Theorem 37.6.4. See Epstein–Petronio [EP94, Theorem 4.14] or Ratcliffe [Rat2006, §11.2, §13.5]; our statement is a special case of the theorem by Maskit [Mas71], but see Remark 37.6.5. □

Remark 37.6.5. The proof of Poincaré's theorem [Poi1882, Poi1883] has a bit of a notorious history. From the very beginning, to quote Epstein–Petronio [EP94, §9, p. 164]:

> It is clear that Poincaré understood very well what was going on. However, the papers are not easy to read. In particular, the reader of the three-dimensional case is referred to the treatment of the two-dimensional case for proofs; this is fully acceptable for a trail-blazing paper, but not satisfactory in the long term.
> There are a number of reasonable published versions of Poincaré's Theorem in dimension two. Of these, we would single out the version by de Rham [dR71] as being particularly careful and easy to read. Most published versions of Poincaré's Theorem applying to all dimensions are unsatisfactory for one reason or another.

This sentiment is echoed by Maskit [Mas71], who presents a proof for polygons with an extension to polyhedron, with the opening remark:

> There are several published proofs of [Poincaré's classical] theorem, but there is some question as to their validity; Siegel [Sie65, p. 115] has commented on this and given an apparently valid proof under fairly restrictive conditions. None of the published proofs are as general as they might be, and they all have a convexity condition that is never really used.
> This note is an attempt to clarify the situation. The problem and the solution presented below arose during the course of several informal conversations. Present at one or more of these conversations were L. V. Ahlfors, L. Bers, W. Magnus, J. E. McMillan, and B. Maskit.

Epstein and Petronio [EP94, §9, p. 165] then have this to say:

> Maskit's paper contains a nice discussion of completeness, though again it is not a constructive approach. He limits his discussion to hyperbolic space in dimensions two and three. We are not confident that the arguments in the paper are complete. For example, there seems to be an assumption that the quotient of a metric space, such that the inverse image of a point is finite, is metric. This is false, as is shown by identifying x with $-x$ in $[-1, 1]$, provided $0 \leq x < 1$.

Maskit published a book [Mas88] containing an expanded version of the proof for polyhedra, to which Epstein and Petronio [EP94, §9, p. 164] review:

> The treatment in [Mas88] is difficult to understand. For example in H.9 on page 75, it is claimed that a metric is defined in a certain way, and this fact is said to be "easy to see", but it seems to us an essential and non-trivial point, which is not so easy to see, particularly when the group generated by the face-pairings is not discrete. . . . The Proposition in IV.1.6 on page 79 of this book is incorrect—a counter-example is given in Example 9.1—because there are no infinite cycles or infinite edges according to the definitions in the book.

37.7 Signature of a Fuchsian group

As an application of Theorem 37.5.3, we relate area and signature for good orbifolds obtained from Fuchsian groups.

37.7.1. We first recall 34.8.13: a good compact, oriented 2-orbifold X is classified up to homeomorphism by its signature $(g; e_1, \dots, e_k)$, where g is the genus of the underlying topological surface and the $e_1, \dots, e_k$ are the orders of the (necessarily cyclic) nontrivial stabilizer groups. This extends to good orbifolds Y with finitely many points removed: we define the signature $(g; e_1, \dots, e_k; \delta)$ where $\delta \geq 0$ is the number of punctures.

Now let $\Gamma \leq \mathrm{PSL}_2(\mathbb{R})$ be a Fuchsian group such that the quotient $Y(\Gamma) = \Gamma\backslash\mathbf{H}^2$ has finite hyperbolic area $\mu(Y(\Gamma))$.

Theorem 37.7.2. (Siegel). *Suppose that $\mu(\Gamma\backslash\mathbf{H}^2)$ has finite hyperbolic area. Then every Dirichlet domain ⌑ has finitely many sides.*

Proof. The proof estimates the contribution to the volume from infinite vertices: for Fuchsian groups, see Siegel [Sie45, p. 716–718], or the expositions of this argument by Gel'fand–Graev–Pyatetskii-Shapiro [GG90, Chapter 1, §1.5] and Katok [Kat92, Theorem 4.1.1]. □

37.7.3. By Theorem 37.5.3 and Siegel's theorem (Theorem 37.7.2), a Dirichlet domain ⌑ for Γ is connected, convex, hyperbolic polygon. In particular, there are $m \in \mathbb{Z}_{\geq 0}$ vertex cycles, which by 37.3.10 correspond to cyclic stabilizer groups of orders $e_1, \dots, e_m \in \mathbb{Z}_{\geq 1}$ listed so that $e_1, \dots, e_k \geq 2$, and finitely many $\delta \in \mathbb{Z}_{\geq 0}$ infinite vertex cycles, corresponding to δ stabilizer groups that are infinite cyclic.

Proposition 37.7.4. *We have*

$$\mu(\Gamma\backslash\mathbf{H}^2) = 2\pi\left((2g-2) + \sum_{i=1}^{k}\left(1 - \frac{1}{e_i}\right) + \delta\right).$$

Proof. Let ⌑ be a Dirichlet domain for Γ with $2n$ sides and n finite or infinite vertices. The hyperbolic area of ⌑ is given by the Gauss–Bonnet formula 33.6.9: summing vertex cycles using the cycle condition (Lemma 37.6.2), we have

$$\mu(⌑) = (2n-2)\pi - \sum_{i=1}^{k} \frac{2\pi}{e_i}.$$

The quotient $\Gamma\backslash\mathbf{H}^2$ is a (punctured) oriented topological surface of genus g, with $k + \delta$ vertices, n edges, and 1 face. By Euler's formula, we have

$$2 - 2g = (k + \delta) - n + 1$$

so

$$n - 1 = 2g - 2 + (k + \delta).$$

Therefore

$$\mu(\text{⌑}) = 2\pi\left((2g-2)+(k+\delta)-\sum_{i=1}^{k}\frac{1}{e_i}\right) = 2\pi\left((2g-2)+\sum_{i=1}^{k}\left(1-\frac{1}{e_i}\right)+\delta\right)$$

as claimed. □

37.8 The (6, 4, 2)-triangle group

We pause to refresh the quaternionic thread and consider a particularly nice example, showing how quaternion algebras arise naturally in the context of Fuchsian groups.

Consider a hyperbolic triangle T with angles $\pi/2, \pi/6, \pi/6$ and vertices placed as in Figure 37.8.1.

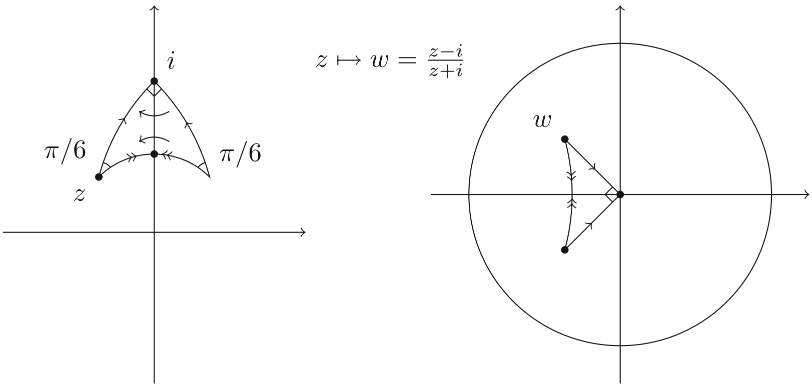


Figure 37.8.1: The hyperbolic triangle T with angles $\pi/2, \pi/6, \pi/6$

By symmetry, we may define the side-pairing P as shown in Figure 37.8.1. This polygon satisfies the cycle condition, so by the Poincaré polygon theorem (Theorem 37.6.4), there exists a Fuchsian group $\Delta \subset \mathrm{PSL}_2(\mathbb{R})$ generated by the two side pairing elements in P and with fundamental domain T. In this section, we construct this group explicitly and observe some interesting arithmetic consequences.

37.8.2. We seek the position of the vertex $z \in \mathbf{H}^2$ corresponding to $w \in \mathbf{D}^2$. Zooming in, we obtain the picture as in Figure 37.8.3.

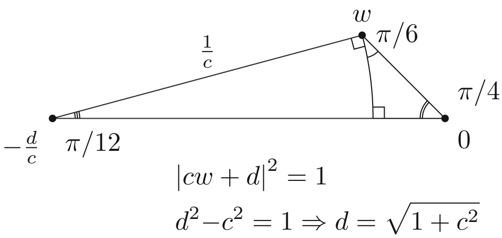


Figure 37.8.3: Finding the position of z

The edge containing w and its complex conjugate is defined by an isometric circle $|cw+d| = 1$ with $d^2 - c^2 = 1$ and by symmetry $c, d \in \mathbb{R}_{>0}$. With the angles as labeled, we find that $w = (-d + e^{\pi i/12})/c$, and since $\arg(w) = 3\pi/4$ we compute that

$$\sqrt{1+c^2} - \cos(\pi/12) = \sin(\pi/12) \tag{37.8.4}$$

so $c^2 = 2\cos(\pi/12)\sin(\pi/12) = \sin(\pi/6) = 1/2$ thus $c = 1/\sqrt{2}$ and $d = \sqrt{3/2}$, so this isometric circle is defined by $|w + \sqrt{3}|^2 = 2$, and $w = (-1+i)/(1+\sqrt{3})$. By coincidence, we find that $z = z_2 = (-1+i)/(1+\sqrt{3}) = w$ as well. The intersection of this circle with the imaginary axis is the point $z_3 = (\sqrt{3}-1)i/\sqrt{2}$.

The unique element mapping the sides meeting at i is obtained by pulling back rotation by $-\pi/4$ in the unit disc model; it is given by the matrix

$$\delta_4 = \begin{pmatrix} 1 & -i \\ 1 & i \end{pmatrix} \begin{pmatrix} e^{-\pi i/4} & 0 \\ 0 & e^{\pi i/4} \end{pmatrix} \begin{pmatrix} 1 & -i \\ 1 & i \end{pmatrix}^{-1} = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 & -1 \\ 1 & 1 \end{pmatrix} \tag{37.8.5}$$

with

$$\delta_4(z) = \frac{z-1}{z+1}, \qquad \text{for } z \in \mathbf{H}^2,$$

and $\delta_4^4 = 1 \in \mathrm{PSL}_2(\mathbb{R})$. From a similar computation, we find that the other side pairing element acting by hyperbolic rotation around z_3 is

$$\delta_2 = \frac{1}{\sqrt{2}} \begin{pmatrix} 0 & -1+\sqrt{3} \\ -1-\sqrt{3} & 0 \end{pmatrix} \tag{37.8.6}$$

and $\delta_2^2 = 1 \in \mathrm{PSL}_2(\mathbb{R})$. We have $(\delta_2\delta_4)(z_4) = \delta_2(z_2) = z_4$, so the element

$$\delta_6 = (\delta_2\delta_4)^{-1} = \frac{1}{2} \begin{pmatrix} 1+\sqrt{3} & 1-\sqrt{3} \\ 1+\sqrt{3} & -1+\sqrt{3} \end{pmatrix} \tag{37.8.7}$$

fixes the vertex z_4, with $\delta_6(z_1) = (1+\sqrt{-3})/2$ and $\delta_6^6 = 1 \in \mathrm{PSL}_2(\mathbb{R})$. These actions are summarized in Figure 37.8.8.

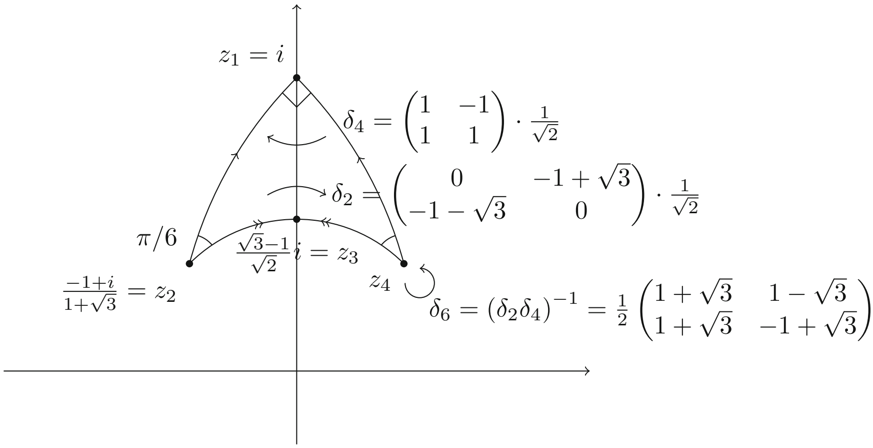


Figure 37.8.8: Side-pairing elements for the (6, 4, 2)-triangle group

So recording these cycle relations,

$$\begin{aligned}\Delta &= \langle \delta_2, \delta_4 \mid \delta_2^2 = \delta_4^4 = (\delta_2\delta_4)^6 = 1\rangle \\ &= \langle \delta_2, \delta_4, \delta_6 \mid \delta_2^2 = \delta_4^4 = \delta_6^6 = \delta_2\delta_4\delta_6 = 1\rangle.\end{aligned}$$

37.8.9. From this setup, we can identify a quaternion algebra obtained from (appropriately scaled) generators of Δ. We have the characteristic polynomials

$$\delta_2^2 + 1 = \delta_4^2 - \sqrt{2}\delta_4 + 1 = \delta_6^2 - \sqrt{3}\delta_6 + 1 = 0.$$

To obtain rational traces, and to simplify the presentation a bit further we consider the $\mathbb{Z}$-subalgebra $B \subseteq \mathrm{M}_2(\mathbb{R})$ generated by

$$i := \sqrt{2}\delta_4, \quad j := \sqrt{3}\delta_6, \quad k := -\sqrt{6}\delta_2;$$

we have

$$\begin{aligned} i^2 &= 2i - 2 & jk &= -3\bar{i} \\ j^2 &= 3j - 3 & ki &= -2\bar{j} \\ k^2 &= -6 & ij &= -\bar{k} \end{aligned} \tag{37.8.10}$$

so as in (22.1.2) with $(a, b, c, u, v, w) = (-3, -2, -1, 2, 3, 0)$, we obtain an order $\mathcal{O} = \mathbb{Z} + \mathbb{Z}i + \mathbb{Z}j + \mathbb{Z}k$ of reduced discriminant $-24 + 12 + 18 = 6$ and with associated ternary quadratic form

$$-3x^2 - 2y^2 - z^2 + 2yz + 3xz.$$

37.8.11. We now try to simplify the presentation (37.8.10) as much as possible. This kind of activity is more aesthetics than mathematics, but for this example there is a preferred way of writing the algebra and order (regrettably, not in a good basis) as follows. We write $i' = i - 1$ and $j' = i(j - 3/2i) = -3i + k + 3$ so that now $(i')^2 = -1$ and $(j')^2 = 3$ and $j'i' = -i'j'$; the remaining basis element of the order in terms of these generators can be taken to be $k' = (1 + i' + j' + i'j')/2 = 3 - i - j + k$.

37.8.12. Throwing away primes from the previous paragraph, we have the algebra $B := \left(\dfrac{-1, 3}{\mathbb{Q}}\right)$ of discriminant $\operatorname{disc} B = 6$ and order

$$\mathcal{O} := \mathbb{Z} \oplus \mathbb{Z}i \oplus \mathbb{Z}j \oplus \mathbb{Z}k, \quad k = \frac{1 + i + j + ij}{2} \tag{37.8.13}$$

with $k^2 - k - 1 = 0$. (By 23.1.1, since $\operatorname{discrd}(\mathcal{O}) = \operatorname{disc} B = 6$, we conclude that $\mathcal{O}$ is a maximal order in B.)

This algebra came with the embedding

$$\begin{aligned} \iota_\infty \colon B &\hookrightarrow \mathrm{M}_2(\mathbb{R}) \\ i, j &\mapsto \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}, \begin{pmatrix} \sqrt{3} & 0 \\ 0 & -\sqrt{3} \end{pmatrix} \\ t + xi + yj + zij &\mapsto \begin{pmatrix} t + y\sqrt{3} & -x + z\sqrt{3} \\ x + z\sqrt{3} & t - y\sqrt{3} \end{pmatrix} = g. \end{aligned}$$

Following the transformations back, we recover the triangle group as a subgroup of $B^\times/F^\times$ after rescaling to be

$$\begin{aligned}\delta_2 &:= 3i + ij = -1 + 2i - j + 2k,\\ \delta_4 &:= 1 + i,\\ \delta_6 &:= 3 + 3i + j + ij = 2 + 2i + 2k;\end{aligned} \tag{37.8.14}$$

and we confirm

$$\delta_2^2 = -6, \quad \delta_4^4 = -4, \quad \delta_6^6 = -1728, \quad \delta_2\delta_4\delta_6 = -12$$

so the images of these elements in $B^\times/F^\times$ generate the group Δ.

In fact, the elements $\delta_2, \delta_4, \delta_6 \in N_{B^\times}(O)$ normalize the order O:

$$\begin{array}{lll}\delta_2^{-1} i\delta_2 = -1 + i - j + 2k & \delta_2^{-1} j\delta_2 = -j & \delta_2^{-1} k\delta_2 = 1 - k\\ \delta_4^{-1} i\delta_4 = i & \delta_4^{-1} j\delta_4 = -ij & \delta_4^{-1} k\delta_4 = 1 + i + j - k\\ \delta_6^{-1} i\delta_6 = -1 + i - j + 2k & \delta_6^{-1} j\delta_6 = 2 - i + 2j - 4k & \delta_6^{-1} k\delta_6 = 1 + j - k\end{array}$$

37.9 Unit group for discriminant 6

In this section, we continue the example from the previous section, specifically the order O in 37.8.12. Let

$$\Gamma := \iota_\infty(O^1/\{\pm 1\}) \leq \mathrm{PSL}_2(\mathbb{R}).$$

By organized enumeration, we will see that Γ is a Fuchsian group and compute a fundamental domain for the action of Γ. We return to this example in section 38.1 as a basic example of the general theory of arithmetic groups arising from quaternion algebras.

37.9.1. Moving to the unit disc via the map $z \mapsto \phi(z) = w = (z - i)/(z + i)$,

$$g^\phi = \begin{pmatrix} t - ix & \sqrt{3}(y - iz) \\ \sqrt{3}(y + iz) & t + ix \end{pmatrix}.$$

To avoid cumbersome notation, we identify g with g^ϕ.

The isometric circle of such an element g is a

$$\text{circle of radius } \frac{1}{\sqrt{3(y^2 + z^2)}} \text{ centered at } \frac{-(t + ix)}{\sqrt{3}(y + iz)}$$

when $y^2 + z^2 \neq 0$; when $y^2 + z^2 = 0$, the center $w = 0 \in \mathbf{D}^2$ is stabilized, with stabilizer $\mathrm{Stab}_\Gamma(0) = \langle S \rangle$ where $S = \begin{pmatrix} i & 0 \\ 0 & -i \end{pmatrix}$ acts by rotation by π. We have

$$\maltese(\Gamma; 0) = \bigcap_{\gamma \in \Gamma - \langle S \rangle} \mathrm{cl}\,\mathrm{ext}\, I(\gamma).$$

To make a fundamental domain, we need to intersect ⌑(0) with a fundamental domain for $\langle S \rangle$, and we take the left half-plane. Then

$$⌑ = ⌑(\Gamma; 0) \cap \{w \in \mathbf{D}^2 : \operatorname{Re} w \le 0\}$$

is a fundamental domain for Γ.

37.9.2. We list elements $g = t + xi + (y + zi)j$ with $\det g = t^2 + x^2 - 3(y^2 + z^2) = 1$ and $t, x, y, z \in \frac{1}{2}\mathbb{Z}$ whose doubles are all of the same parity. We enumerate them in increasing order of the (square) inverse radius $y^2 + z^2$ (ignoring the factor 3). The case $y^2 + z^2 = 0$ gives us the stabilizer. By parity, we cannot have $y^2 + z^2 = 1/4$. If $y^2 + z^2 = 1/2$ then we find $y, z = \pm 1/2$ and $t^2 + x^2 = 5/2$. Sifting out all of the possibilities, we find

$$g = \pm\frac{3}{2} \pm \frac{1}{2}i \pm \frac{1}{2}j \pm \frac{1}{2}ij \quad \text{or} \quad g = \pm\frac{1}{2} \pm \frac{3}{2}i \pm \frac{1}{2}j \pm \frac{1}{2}ij.$$

All of these elements have the radius of $I(g)$ equal to $\sqrt{2/3} = 0.82\ldots$, and the centers are $\pm 1.15 \pm 0.57\sqrt{-1}$ and $\pm 0.57 \pm 1.15\sqrt{-1}$. The corresponding external domain is given in Figure 37.9.3.

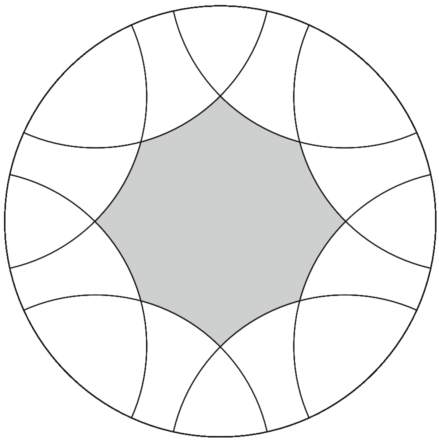


Figure 37.9.3: A starting external domain

37.9.4. If we continue in this way, listing elements according to decreasing radius, we find that all remaining elements have too small a radius to cut away anything extra from the external domain. The corresponding external domain looks like Figure 37.9.5.

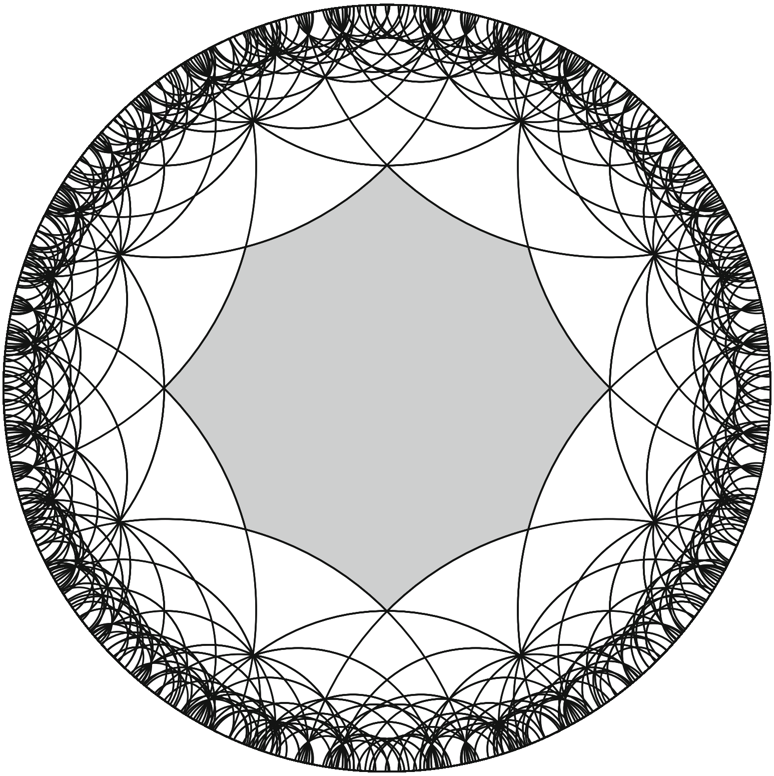


Figure 37.9.5: The external domain

It follows that the ⌑ is cut out by the left half with side pairing as in Figure 37.9.6.

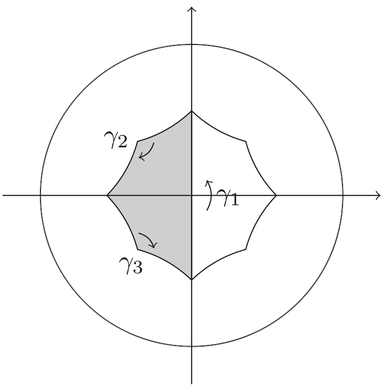


Figure 37.9.6: ⌑ $\subset \mathbf{D}^2$

Pulling back to the upper half-plane, we obtain the domain in Figure 37.9.7.

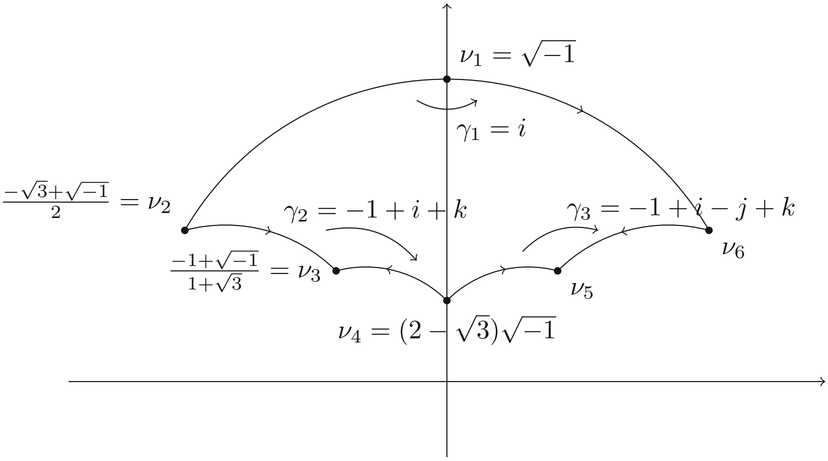


Figure 37.9.7: ⌑ $\subset \mathbf{H}^2$

The corresponding tessellation of the upper half-plane looks like the one in Figure 37.9.8.

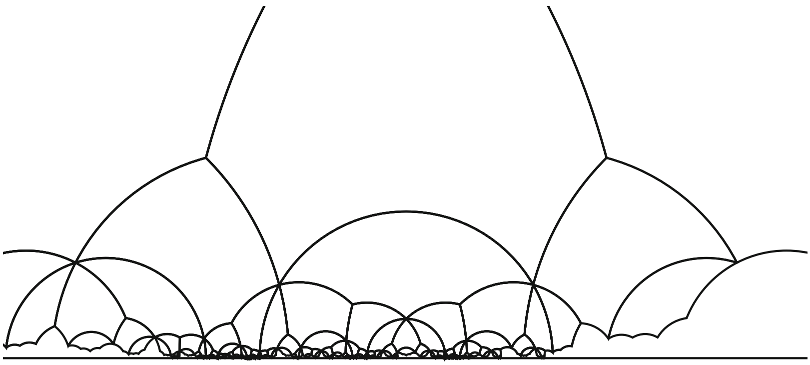


Figure 37.9.8: Tessellation of $\mathbf{H}^2$ by Γ

In addition to this side pairing, we check the cycle relations on the four orbits of vertices: the fixed points v_1, v_3, v_5 and the orbit v_2, v_4, v_6. In conclusion,

$$\Gamma = \langle \gamma_1, \gamma_2, \gamma_3 \mid \gamma_1^2 = \gamma_2^3 = \gamma_3^3 = (\gamma_3\gamma_2\gamma_1)^2 = 1 \rangle \tag{37.9.9}$$

and letting $\gamma_4 = \gamma_3\gamma_2\gamma_1 = -2i + j$, we can rewrite this more symmetrically as

$$\Gamma = \langle \gamma_1, \gamma_2, \gamma_3, \gamma_4 \mid \gamma_1^2 = \gamma_2^3 = \gamma_3^3 = \gamma_4^2 = \gamma_4\gamma_3\gamma_2\gamma_1 = 1 \rangle.$$

The order of the stabilizers tell us the angles at each vertex, and so by the Gauss–Bonnet formula (Theorem 33.6.8) we compute that the area is

$$\mu(⌑) = 3\pi - (2/3 + 2/3 + 1)\pi = 2\pi/3.$$

37.9.10. Finally, the quotient $X(\Gamma) = \Gamma\backslash\mathbf{H}^2$ has the structure of a good complex 1-orbifold (see 34.8.14), a Riemann surface but with finitely many orbifold points. Since the fundamental domain ⌑ is compact, via the continuous surjective projection map ⌑ $\to X(\Gamma)$ we see that $X(\Gamma)$ is compact, and $\mu(X(\Gamma)) = \mu(⌑) = 2\pi/3$. This orbifold 'folds' up to a surface with topological genus 0, so the signature 34.8.13 of $X(\Gamma)$ is $(0; 2, 2, 3, 3)$: see Figure 37.9.11.

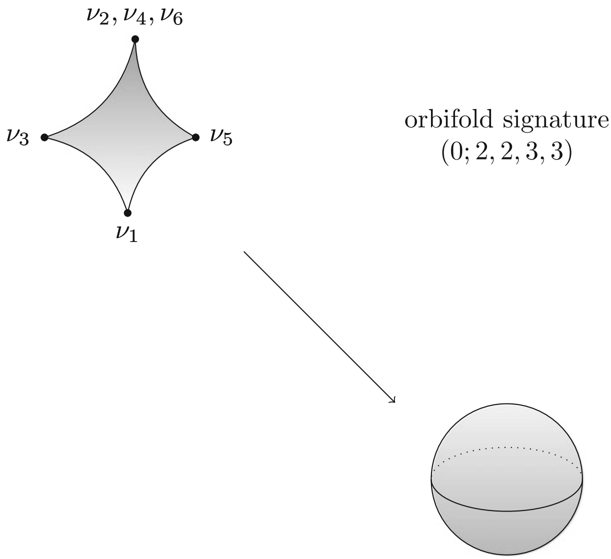


Figure 37.9.11: $X(\Gamma)$ as an orbifold and Riemann surface

We have seen that the norm 1 group contains the $(2, 4, 6)$-triangle group Δ, so we have a map $X(\Gamma) \to X(\Delta) = \Delta\backslash\mathbf{H}^2$; by Gauss–Bonnet, we have

$$\mu(X(\Delta)) = 2(\pi - (1/2 - 1/4 - 1/6)\pi) = \pi/6,$$

we see

$$[\Delta : \Gamma] = \mu(X(\Gamma))/\mu(X(\Delta)) = (2\pi/3)/(\pi/6) = 4,$$

as is visible from Figure 37.9.12 for $\Gamma \overset{4}{\leq} \Delta$.

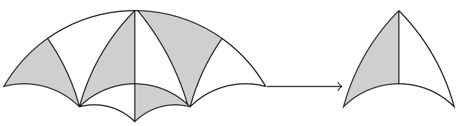


Figure 37.9.12: Triangulation of ⌑

Remark 37.9.13. The discriminant 6 quaternion algebra has been a favorite to study, and the fundamental domain described above is also given by e.g. Alsina–Bayer

[AB2004, §5.5.2] and Kohel–Verrill [KV2003, §5.1]. We return to this example in section 43.2 in the context of abelian surfaces with quaternionic multiplication.

Exercises

1. Let $\Gamma = \mathrm{PSL}_2(\mathbb{Z})$. Describe the Dirichlet domain $\text{⌑}(z)$ for an arbitrary $z \in \mathbf{H}^2$ with $\operatorname{Im} z > 1$.
2. Let $\Gamma = \mathrm{PSL}_2(\mathbb{Z}[i])$ (cf. section 36.6). Show that

$$\text{⌑}(\Gamma; 2j) = \{z = x + yj \in \mathbf{H}^3 : |\mathrm{Re}\, x|, |\mathrm{Im}\, x| \leq 1/2 \text{ and } \|z\| \geq 1\}$$

and that

$$\mathrm{Stab}_\Gamma(2j) = \left\langle \begin{pmatrix} i & 0 \\ 0 & -i \end{pmatrix} \right\rangle \simeq \mathbb{Z}/2\mathbb{Z}$$

so $\text{⌑}(\Gamma; 2j)$ is a union of two copies of a fundamental set for Γ.

3. Let Γ be the cyclic Fuchsian group generated by the isometry $z \mapsto 4z$, represented by $\begin{pmatrix} 2 & 0 \\ 0 & 1/2 \end{pmatrix} \in \mathrm{PSL}_2(\mathbb{R})$. Give an explicit description of the Dirichlet domain $\text{⌑}(\Gamma; i) \subset \mathbf{H}^2$ and its image $\text{⌑}(\Gamma; 0) \subset \mathbf{D}^2$ with $i \mapsto 0$.
4. Let $g = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in \mathrm{PSL}_2(\mathbb{R})$ satisfy $gi \neq i$. Let H be the perpendicular bisector between i and gi.

   (a) Show that $\|g\|^2 > 2$.
   (b) Under the map $\mathbf{H}^2 \to \mathbf{D}^2$ taking $i \mapsto 0$, show that the perpendicular bisector between i and gi is the isometric circle $I(g^{-1})$ inside $\mathbf{D}^2$.
   (c) Show that H is the half-circle of square radius $\dfrac{\|g\|^2 - 2}{(c^2 + d^2 - 1)^2}$ centered at $\dfrac{ac + bd}{c^2 + d^2 - 1} \in \mathbb{R}$, where $\|g\|^2 = a^2 + b^2 + c^2 + d^2$. *[Hint: as a check along the way, H is described by the equation $|(d - ic)z + i(a + ib)| = |z + i|$.]*

5. Let

$$g = \begin{pmatrix} \overline{d} & \overline{c} \\ c & d \end{pmatrix} \in \mathrm{PSL}(1,1) \circlearrowright \mathbf{D}^2$$

with $c, d \in \mathbb{C}$ satisfying $|d|^2 - |c|^2 = 1$. Show directly that $|gw| = |w|$ for $w \in \mathbf{D}^2$ if and only if

$$|cw + d| = 1$$

by expanding out and simplifying.

6. Show that for all $g \in \mathrm{PSU}(1,1)$, we have $gI(g) = I(g^{-1})$, where $I(g)$ is the isometric circle of g. (Equivalently, show that if $g \in \mathrm{PSL}_2(\mathbb{R})$ that $gL(g; z_0) = L(g^{-1}; z_0)$ for all $z_0 \in \mathbf{H}^2$.)
7. ▷ Let ⌑ be a Dirichlet domain for a Fuchsian group Γ. Show that in every compact set, there are only finitely many sides and finitely many vertices of ⌑.

8. Let Γ be a discrete group of isometries acting properly on a locally compact, complete metric space X, and let $x_0 \in X$. Recall the definition of $H(\gamma; x_0)$ (37.4.6) for $\gamma \in \Gamma$ and $L(\gamma; z_0) = \operatorname{bd} H(\gamma; z_0)$. Show that if K is a compact set, then $K \cap L(\gamma; x_0) \neq \emptyset$ for only finitely many $\gamma \in \Gamma$.

9. Extend Theorem 37.4.18 as follows. Let (X, ρ) be a complete, locally compact geodesic space, let Γ be a discrete group of isometries acting properly on X. Let $x_0 \in X$, let $\Gamma_0 = \operatorname{Stab}_\Gamma(x_0)$, and let $u_0 \in X$ be such that $\operatorname{Stab}_{\Gamma_0}(u_0) = \{1\}$. Show that Γ_0 is a discrete group of isometries acting properly on X, and

$$⌑(\Gamma; z_0) \cap ⌑(\Gamma_0; u_0)$$

is a locally finite fundamental set for Γ that is star-shaped with respect to x_0 and whose boundary consists of hyperplane bisectors.

10. Consider the *egg of revolution*, a surface of revolution obtained from convex curves with positive curvature as in Figure 37.9.14.

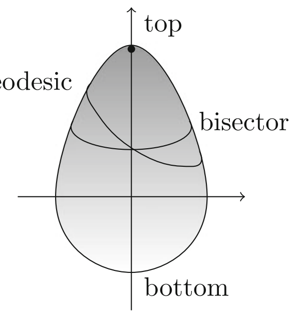


Figure 37.9.14: Egg of revolution

An egg of revolution has the structure of a geodesic space with the induced metric from $\mathbb{R}^3$. Show that the separator between the top and bottom, a circle of revolution, is *not* geodesic. [In fact, Clairaut's relation shows that the geodesic joining two points in the same circle of revolution above crest in the x-axis never lies in this circle of revolution.]

11. In this exercise, we consider a Fuchsian group constructed from a regular quadrilateral.

    (a) Show that for every $e \geq 2$, there exists a regular (equilateral and equiangular) quadrilateral $⌑ \subset \mathbf{D}^2$, unique up to isometry, with interior angle $\pi/(2e)$.

    Conclude from Poincaré's theorem that there is a Fuchsian group, unique up to conjugation in $\mathrm{PSL}_2(\mathbb{R})$, with fundamental domain $⌑$ and side pairing

as in Figure 37.9.15.

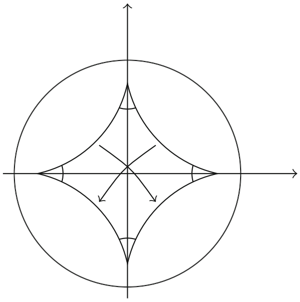


Figure 37.9.15: A special quadrilateral Fuchsian group

(b) Give a presentation for this group for all $e \geq 2$, and find explicit matrix generators for the special case when $e = 2$.

This chapter is licensed under the terms of the Creative Commons Attribution-NonCommercial 4.0 International License (http://creativecommons.org/licenses/by-nc/4.0/), which permits any noncommercial use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license and indicate if changes were made.

The images or other third party material in this chapter are included in the chapter's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the chapter's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.

Chapter 38

Quaternionic arithmetic groups

In this chapter, we now apply our topological and geometric interpretation of discrete groups in our case of interest: quaternion unit groups.

38.1 ▹ Rational quaternion groups

The classical modular group $\mathrm{PSL}_2(\mathbb{Z})$ (chapter 35) was obtained as follows: we

- started with the matrix algebra $\mathrm{M}_2(\mathbb{Q})$,
- looked inside for integral elements to find the order $\mathrm{M}_2(\mathbb{Z})$,
- took its unit group $\mathrm{GL}_2(\mathbb{Z})$, and finally
- restricted attention to the Fuchsian group $\mathrm{PSL}_2(\mathbb{Z})$ to get a faithful action on the upper half-plane by orientation-preserving isometries.

The wonderful thing that gives life to this part of our monograph is this: the same thing works if we replace $\mathrm{M}_2(\mathbb{Q})$ with a quaternion algebra B over a global field! In this section, we derive key aspects of this program for quaternion algebras over $\mathbb{Q}$ in a self-contained way, before embarking on a more general study in the remainder of the chapter.

In section 32.1, we already dealt with the case when the quaternion algebra was definite, finding a finite unit group; so here we take $B = \left(\dfrac{a,b}{\mathbb{Q}}\right)$ indefinite. Without loss of generality (for convenience of presentation), we may further suppose $a, b > 0$ and both $a, b \in \mathbb{Z}$. To be indefinite means that $B \otimes_{\mathbb{Q}} \mathbb{R} = \left(\dfrac{a,b}{\mathbb{R}}\right) \simeq \mathrm{M}_2(\mathbb{R})$; we obtain such an embedding via a conjugate of the left-regular representation (2.3.2)

$$\begin{aligned} \iota_\infty \colon B &\hookrightarrow \mathrm{M}_2(\mathbb{R}) \\ t + xi + yj + zk &\mapsto \begin{pmatrix} t + x\sqrt{a} & (y + z\sqrt{a})\sqrt{b} \\ (y - z\sqrt{a})\sqrt{b} & t - x\sqrt{a}. \end{pmatrix} \end{aligned} \tag{38.1.1}$$

This embedding is not unique, but another embedding would correspond to post-composition by an $\mathbb{R}$-algebra automorphism of $\mathrm{M}_2(\mathbb{R})$, which by the Skolem–Noether

© The Author(s) 2021

J. Voight, *Quaternion Algebras*, Graduate Texts in Mathematics 288,
https://doi.org/10.1007/978-3-030-56694-4_38

theorem (Corollary 7.1.4) is given by conjugation by an element of $\mathrm{GL}_2(\mathbb{R})$, and so we can live with a choice and this ambiguity.

Let
$$\mathcal{O} := \mathbb{Z}\langle i, j\rangle = \mathbb{Z} + \mathbb{Z}i + \mathbb{Z}j + \mathbb{Z}k$$
then $\mathcal{O} \subset B$ is an order, and $\iota_\infty(\mathcal{O}) \subseteq \mathrm{M}_2(\mathbb{Z})$ consists of the subset of matrices in (38.1.1) with $t, x, y, z \in \mathbb{Z}$. Following the script above, we define
$$\mathcal{O}^1 := \{\gamma \in \mathcal{O}^\times : \mathrm{nrd}(\gamma) = 1\}$$
and let
$$\Gamma^1(\mathcal{O}) := \iota(\mathcal{O}^1)/\{\pm 1\} \subseteq \mathrm{PSL}_2(\mathbb{R}).$$

Lemma 38.1.2. *The group* $\Gamma^1(\mathcal{O}) \leq \mathrm{PSL}_2(\mathbb{R})$ *is a Fuchsian group.*

Recall that a Fuchsian group (Definition 34.7.3) is a discrete subgroup of $\mathrm{PSL}_2(\mathbb{R})$; by Theorem 34.2.1, a Fuchsian group acts properly on the upper half-plane $\mathbf{H}^2$ by orientation-preserving isometries.

Proof. Because $\Gamma^1(\mathcal{O})$ is a group, it suffices to find an open neighborhood of 1 containing no other element of $\Gamma^1(\mathcal{O})$. We take
$$U = \left\{ \pm \begin{pmatrix} g_{11} & g_{12} \\ g_{21} & g_{22} \end{pmatrix} \in \mathrm{PSL}_2(\mathbb{R}) : |g_{11} - 1|, |g_{12}|, |g_{21}|, |g_{22} - 1| < 1/2 \right\}.$$
If $\gamma = (g_{ij})_{i,j} \in U \cap \Gamma^1(\mathcal{O})$, then
$$\begin{aligned} |2(t-1)| = |g_{11} + g_{22} - 2| &\leq |g_{11} - 1| + |g_{22} - 1| < 1 \\ |2y\sqrt{b}| = |g_{12} + g_{21}| &\leq |g_{12}| + |g_{21}| < 1 \end{aligned}$$
and since $t, y \in \mathbb{Z}$ we have $t = 1$ and $y = 0$. Then
$$\begin{aligned} |x\sqrt{a}| &= |g_{11}| < 1/2 \\ |z\sqrt{ab}| &= |g_{12}| < 1/2 \end{aligned}$$
and since $a, b, x, z \in \mathbb{Z}$ with $a, b \neq 0$, we conclude $x = z = 0$, and $\gamma = \pm 1$. $\square$

If $B \simeq \mathrm{M}_2(\mathbb{Q})$, then $\Gamma^1(\mathcal{O}) = \mathrm{PSL}_2(\mathbb{Z})$ and we investigated this case already in detail. So suppose from now on that $B \not\simeq \mathrm{M}_2(\mathbb{Q})$, or equivalently that B is a division algebra over $\mathbb{Q}$.

Proposition 38.1.3. *The quotient* $\Gamma^1(\mathcal{O})\backslash\mathbf{H}^2$ *is compact.*

This proposition is analogous to the finiteness of the class set (as in section 17.5), and the proof is again inspired by the geometry of numbers. We give a proof in Main Theorem 38.4.3.

The compactness result above implies nice properties for $\Gamma^1(\mathcal{O})$, read off from a fundamental domain (as presented in section 37.1). Let $z_0 \in \mathbf{H}^2$ have trivial stabilizer $\mathrm{Stab}_{\Gamma^1(\mathcal{O})} z_0 = \{1\}$. Then the Dirichlet domain
$$\text{⌑} = \text{⌑}(\Gamma^1(\mathcal{O}); z_0) = \{z \in \mathbf{H}^2 : \rho(z, z_0) \leq \rho(\gamma z, z_0) \text{ for all } \gamma \in \Gamma^1(\mathcal{O})\}$$

is a closed, locally finite fundamental domain for $\Gamma^1(O)$ with geodesic sides by Theorem 37.1.12.

Corollary 38.1.4. *⌑ is a compact, finite-sided hyperbolic polygon, and the group* $\Gamma^1(O)$ *is finitely presented.*

Proof. We write $\Gamma = \Gamma^1(O)$. Since $\Gamma\backslash\mathbf{H}^2$ is compact by Proposition 38.1.3, the distance $\rho(\Gamma z_0, \Gamma z)$ for $\Gamma z \in X(\Gamma)$ is bounded. Thus by construction, the Dirichlet domain is contained in a bounded set and is therefore compact. Since ⌑ is locally finite, we conclude that γ⌑ $\cap$ ⌑ $\neq \emptyset$ for only finitely many $\gamma \in \Gamma$. But the set of such elements are the side pairing elements and they generate Γ (Theorem 37.3.1), so ⌑ has finitely many sides and Γ is finitely generated. Thus ⌑ has finitely many vertices, the set of vertex cycle relations is finite, and these generate the relations (Proposition 37.3.14). Thus Γ is finitely presented. □

Two subgroups $H_1, H_2 \leq G$ are **commensurable** if $H_1 \cap H_2$ has finite index in both H_1, H_2.

38.1.5. If $\Gamma \leq \mathrm{PSL}_2(\mathbb{R})$ is commensurable with $\Gamma^1(O)$, we say that Γ is an **arithmetic** Fuchsian group with defining quaternion algebra B. This definition is independent of the choice of order O: another suborder or superorder has finite index, so the corresponding unit groups will also have finite index.

For every Fuchsian group $\Gamma \leq \mathrm{PSL}_2(\mathbb{R})$ commensurable with $\Gamma^1(O)$, the conclusions of Proposition 38.1.3 and Corollary 38.1.4 remain true: for a containment $\Gamma' \leq \Gamma$ of finite index, the corresponding map $\Gamma'\backslash\mathbf{H}^2 \to \Gamma\backslash\mathbf{H}^2$ is finite-to-one, and the desired properties pass from one quotient to the other.

In this way, we have completed our task: starting with an indefinite quaternion algebra B over $\mathbb{Q}$, we constructed a Fuchsian group $\Gamma^1(O)$ generalizing $\mathrm{PSL}_2(\mathbb{Z})$ acting on the upper half-plane. We pursue a more general construction in this chapter (considering an indefinite quaternion over a number field) and its geometry and arithmetic properties in the remainder of this book.

38.2 Isometries from quaternionic groups

In the remainder of this chapter, we investigate discrete groups obtained from unit groups of quaternion algebra over number fields. Throughout, let F be a number field with r real places and c complex places, so that $[F : \mathbb{Q}] = r + 2c = n$. Let B be a quaternion algebra over F.

38.2.1. Suppose that B is split at t real places and ramified at the remaining $r - t$ real places, so that

$$B \hookrightarrow B_{\mathbb{R}} := B \otimes_{\mathbb{Q}} \mathbb{R} \xrightarrow{\sim} \mathrm{M}_2(\mathbb{R})^t \times \mathbb{H}^{r-t} \times \mathrm{M}_2(\mathbb{C})^c. \tag{38.2.2}$$

Let

$$\iota = (\iota_1, \dots, \iota_{t+c})\colon B \to \mathrm{M}_2(\mathbb{R})^t \times \mathrm{M}_2(\mathbb{C})^c \tag{38.2.3}$$

denote the map (38.2.2) composed with the projection onto the matrix ring factors. Then as long as $t + c > 0$, we have

$$\iota(B^\times) \subset \mathrm{GL}_2(\mathbb{R})^t \times \mathrm{GL}_2(\mathbb{C})^c. \tag{38.2.4}$$

We have $t = c = 0$ if and only if F is totally real and B is totally definite; in this case, the geometry disappears, and we are reduced to considering finite unit groups (see section 32.3). There is still more to say for this case, and we will return to it in chapter 41 in the context of modular forms. But in this part of the text we are following geometric threads, so

$$\text{suppose from now on that } t + c > 0; \tag{38.2.5}$$

this is another way of saying that B is indefinite or equivalently satisfies the Eichler condition (Definition 28.5.1).

38.2.6. We now restrict to a subgroup acting faithfully via orientation-preserving isometries. Recall that

$$F_{>0}^\times = \{x \in F^\times : v(x) > 0 \text{ for all real places } v\}$$

is the group of totally positive elements of F. Let

$$B_{>0}^\times := \{\alpha \in B^\times : \mathrm{nrd}(\alpha) \in F_{>0}^\times\} \tag{38.2.7}$$

be the group of units of B of totally positive reduced norm. Then $F^\times \subset B_{>0}^\times$ because $\mathrm{nrd}(a) = a^2 \in F_{>0}^\times$ for all $a \in F^\times$. Let

$$\mathrm{P}B_{>0}^\times := B_{>0}^\times / Z(B_{>0}^\times) = B_{>0}^\times / F^\times \tag{38.2.8}$$

be the quotient by the center; then ι induces an inclusion

$$\mathrm{P}\iota(B_{>0}^\times) \subset \mathrm{PSL}_2(\mathbb{R})^t \times \mathrm{PSL}_2(\mathbb{C})^c$$

(we have $\mathrm{PSL}_2(\mathbb{C}) \simeq \mathrm{PGL}_2(\mathbb{C})$ and $\mathrm{PGL}_2^+(\mathbb{R}) \simeq \mathrm{PSL}_2(\mathbb{R})$, rescaling by the determinant). Therefore, the group $\mathrm{P}\iota(B_{>0}^\times)$ acts on

$$\mathcal{H} := (\mathbf{H}^2)^t \times (\mathbf{H}^3)^c \tag{38.2.9}$$

on the left faithfully by linear fractional transformations as orientation-preserving isometries (Theorems 33.3.14 and 36.2.14), that is to say

$$\mathrm{P}\iota(B_{>0}^\times) \subset \mathrm{Isom}^+(\mathcal{H}). \tag{38.2.10}$$

Remark 38.2.11. One can equally well consider $\mathrm{P}\iota(B^\times) \subseteq \mathrm{Isom}(\mathcal{H})$. Indeed,

$$\mathrm{PSL}_2(\mathbb{R})^t \times \mathrm{PSL}_2(\mathbb{C})^c \simeq \mathrm{Isom}^+(\mathbf{H}^2)^t \times \mathrm{Isom}^+(\mathbf{H}^3)^c \leq \mathrm{Isom}^+(\mathcal{H})$$

is the subgroup of isometries that preserve each factor and preserving orientation. The full group $\mathrm{Isom}(\mathcal{H})$ includes more: permuting factors of the same kind is an isometry, and with these we have an isomorphism of groups

$$\mathrm{Isom}(\mathcal{H}) \simeq (\mathrm{Isom}(\mathbf{H}^2)^t \times \mathrm{Isom}(\mathbf{H}^3)^c) \rtimes (S_t \times S_c) \tag{38.2.12}$$

where S_m denotes the symmetric group. Since $\mathrm{PSL}_2(\mathbb{R}) \simeq \mathrm{Isom}^+(\mathbf{H}^2) \leq \mathrm{Isom}(\mathbf{H}^2)$ has index 2 and the same for $\mathbf{H}^3$, we conclude that

$$[\mathrm{Isom}(\mathcal{H}) : \mathrm{PSL}_2(\mathbb{R})^t \times \mathrm{PSL}_2(\mathbb{C})^c] = 2^{t+c} t!c!.$$

The orientation-preserving subgroup $\mathrm{Isom}^+(\mathcal{H}) \leq \mathrm{Isom}(\mathcal{H})$ has index 2, containing for example elements which reverse orientation in two components and preserve orientation in the others; if the orientation is reversed in two components $\mathbf{H}^2$, then the resulting isometry is orientation-preserving but not holomorphic in these components.

38.3 Discreteness

We now seek *discrete* subgroups, as follows.

38.3.1. Let $R = \mathbb{Z}_F$ be the ring of integers of F and let $O \subset B$ be an R-order. Let

$$O^1 := \{\gamma \in O^\times : \mathrm{nrd}(\gamma) = 1\} \leq O^\times \tag{38.3.2}$$

be the subgroup of units of reduced norm 1, and let

$$\begin{aligned} \mathrm{P}O^1 &:= O^1/Z(O^1) = O^1/\{\pm 1\} \\ \Gamma^1(O) &:= \mathrm{P}\iota(O^1) \end{aligned} \tag{38.3.3}$$

Then by (38.2.10), we have $\Gamma^1(O) < \mathrm{Isom}^+(\mathcal{H})$.

Definition 38.3.4. A subgroup $\Gamma \leq \mathrm{Isom}^+(\mathcal{H})$ is **arithmetic** if Γ is commensurable with $\Gamma^1(O)$ for a quaternion algebra B and an order $O \subseteq B$ (with respect to some embedding ι).

If Γ is commensurable with $\Gamma^1(O)$ for an order O then it is commensurable with $\Gamma^1(O')$ for every other order O', since every two orders have finite R-index, thus finite index—so we could equally well compare to one fixed (e.g. maximal) order. The class of arithmetic groups contains in particular the quaternionic unit groups with which we started, but contains other groups of interest (including subgroups and discrete supergroups with finite index).

Remark 38.3.5. There is a more general definition of arithmetic group which reduces to this one; see section 38.5.

Right away, we show that arithmetic groups are discrete. To do so, we will need two short lemmas.

Lemma 38.3.6. *Let K, X be Hausdorff topological spaces, and suppose K is compact. Let $\pi\colon K \times X \to X$ be the projection, and let $Y \subseteq K \times X$ be discrete and closed. Then $\pi(Y) \subseteq X$ is discrete.*

Proof. First, Y has no limit points: a limit point of Y would belong to Y, but then Y is discrete so every point is isolated point. For the same reason, every subset of Y is also (discrete and) closed: a limit point of the subset would be a limit point of Y.

Now let $x \in \pi(Y)$ and let $Y_x = Y \smallsetminus \pi^{-1}(x)$. Then $Y_x \subseteq K \times X$ is closed. The set $(K \times X) \smallsetminus Y_x$ is open and contains $K \times \{x\}$, so by the tube lemma, it contains an open set $K \times U$. Then $U \ni x$ is the desired neighborhood. ∎

Lemma 38.3.7. *Let G be a Hausdorff topological group and let $H \le G$ be a discrete subgroup. Then H is closed.*

Proof. Since H is discrete, there is a neighborhood $U \ni 1$ such that $U \cap H = \{1\}$. Further, there exists a neighborhood $V \subseteq U$ such that $V^{-1}V \subseteq U$ (multiplication and inversion are continuous, see Exercise 12.4).

We show $G \smallsetminus H$ is open. For $x \in G$, we have $xV \ni x$ an open neighborhood, and if $h, h' \in xV \cap H$ then $x^{-1}h, x^{-1}h' \in V$ and so $(x^{-1}h)^{-1}(x^{-1}h') = h^{-1}h' \in V^{-1}V \subseteq U$. Therefore $h^{-1}h' = 1$ by the hypothesis on U, so $h = h'$. Thus xV contains at most one element of H. Since G is Hausdorff, when $x \notin H$ we can shrink V if necessary to get $xV \cap H = \emptyset$, as desired. □

Proposition 38.3.8. *Let $\Gamma \le \operatorname{Isom}^+(\mathcal{H})$ be an arithmetic subgroup. Then Γ is discrete.*

Proof. It is enough to prove the proposition for $\Gamma = \Gamma^1(\mathcal{O})$, as discreteness is preserved between commensurable groups (having finite index in their intersection).

The image $\mathcal{O} \hookrightarrow B_{\mathbb{R}}$ as in (38.2.2) is discrete by 17.7.6: we argued using coordinates and noted that $R = \mathbb{Z}_F \hookrightarrow F_{\mathbb{R}}$ is discrete. Therefore the image

$$\mathcal{O}^1 \hookrightarrow B^1_{\mathbb{R}} \simeq (\mathbb{H}^1)^{r-t} \times \mathrm{SL}_2(\mathbb{R})^t \times \mathrm{SL}_2(\mathbb{C})^c \tag{38.3.9}$$

is discrete (by restriction). Further, since $\mathbb{H}^1$ is compact, by Lemmas 38.3.6 and 38.3.7, the image of $\mathcal{O}^1 \hookrightarrow \mathrm{SL}_2(\mathbb{R})^t \times \mathrm{SL}_2(\mathbb{C})^c$ under the projection is discrete. (Any further projection turns out *not* to be discrete; see Exercise 38.2.) □

38.3.10. The group $\Gamma^1(\mathcal{O})$ is a Fuchsian group if and only if $t = 1$ and $c = 0$, i.e. F is totally real and B is ramified at all but one real place; $\Gamma^1(\mathcal{O})$ is a Kleinian group if and only if $t = 0$ and $c = 1$, i.e. F has exactly one complex place and B is ramified at all real places.

Just as for Fuchsian and Kleinian groups, discrete groups admit several equivalent characterizations as follows. (For the notion of a good orbifold, see Definition 34.8.10.)

Proposition 38.3.11. *Let $\Gamma \le \operatorname{Isom}^+(\mathcal{H})$ be a subgroup. Then the following are equivalent:*

(i) Γ *is discrete (with the subspace topology);*
(ii) *For all $z \in \mathcal{H}$, we have $\#\operatorname{Stab}_\Gamma(z) < \infty$ and there exists an open neigborhood $U \ni z$ such that $\gamma U \cap U \neq \emptyset$ implies $\gamma \in \operatorname{Stab}_\Gamma(z)$;*
(iii) *For all compact subsets $K \subseteq \mathcal{H}$, we have $K \cap \gamma K \neq \emptyset$ for only finitely many $\gamma \in \Gamma$; and*
(iv) *For all $z \in \mathcal{H}$, the orbit $\Gamma z \subseteq \mathcal{H}$ is discrete and $\#\operatorname{Stab}_\Gamma(z) < \infty$.*

Moreover, if these equivalent conditions hold, then the quotient $\Gamma\backslash\mathcal{H}$ has the structure of a good Riemann orbifold of dimension $m = 2t + 3c$, and the quotient map

$$\pi \colon \mathcal{H} \to \Gamma\backslash\mathcal{H}$$

is a local isometry at all points $z \in \mathcal{H}$ with $\operatorname{Stab}_\Gamma(z) = \{1\}$.

Proof. We proved this statement in Propositions 34.7.2 and 36.4.1 when $t = 0$ or $c = 0$; the general case follows similarly. □

38.4 Compactness and finite generation

We now consider further properties of arithmetic groups. Refreshing our notation, let $\Gamma \leq \mathrm{Isom}^+(\mathcal{H})$ be an arithmetic group arising from a quaternion algebra B.

38.4.1. The arithmetic discrete groups arising from the case $B = \mathrm{M}_2(F)$ of the matrix ring are of particular interest: they include the case $F = \mathbb{Q}$ giving rise to the classical modular group $\Gamma = \mathrm{PSL}_2(\mathbb{Z})$ studied in chapter 35 as well as the case $F = \mathbb{Q}(i)$ giving rise to the Picard group $\Gamma = \mathrm{PSL}_2(\mathbb{Z}[i])$ examined in section 36.6 and more generally the Bianchi groups. Let $B = \mathrm{M}_2(F)$. Then necessarily $r = t$ (the matrix ring is already split!), so $\mathcal{H} = (\mathbf{H}^2)^r \times (\mathbf{H}^3)^c$, and the embedding ι in (38.2.3) has the simple description

$$\begin{aligned} \iota\colon\ \mathrm{M}_2(F) &\hookrightarrow \mathrm{M}_2(F)_{\mathbb{R}} \simeq \mathrm{M}_2(\mathbb{R})^r \times \mathrm{M}_2(\mathbb{C})^c \\ \alpha = \begin{pmatrix} a & b \\ c & d \end{pmatrix} &\mapsto \left(\begin{pmatrix} a_v & b_v \\ c_v & d_v \end{pmatrix} \right)_v \end{aligned} \tag{38.4.2}$$

where we embed matrices componentwise, indexed by the archimedean places of F.

Much is written about the compactification of $Y(\Gamma)$, and unfortunately it would take us too far afield to fully treat this important topic: the groups so obtained behave differently in several respects than the case where B is a division algebra.

We suppose throughout this section that B is a division algebra.
Let $X(\Gamma) := \Gamma\backslash\mathcal{H}$ be the quotient, a good Riemann orbifold.

Main Theorem 38.4.3 (Hey). *The orbifold $X(\Gamma)$ is compact.*

Proof. We follow Zassenhaus [Zas72, §1]; see also Kleinert [Klt2000, Theorem 1.1].

First, the group Γ is commensurable with $\iota(\mathcal{O}^1)$ for an R-order $\mathcal{O}$; by comparison under maps of finite index, we may suppose $\Gamma = \Gamma^1(\mathcal{O})$.

Second, we claim that it suffices to show that $\mathcal{O}^1\backslash B^1_{\mathbb{R}}$ is compact. Indeed, recall (38.3.9) that

$$\mathcal{O}^1 \hookrightarrow B^1_{\mathbb{R}} \simeq (\mathbb{H}^1)^{r-t} \times \mathrm{SL}_2(\mathbb{R})^t \times \mathrm{SL}_2(\mathbb{C})^c.$$

From the symmetric space models (34.6.3) and (36.3.14), we have homeomorphisms

$$\begin{aligned} \mathrm{SL}_2(\mathbb{R})/\,\mathrm{SO}(2) &\xrightarrow{\sim} \mathbf{H}^2 \\ \mathrm{SL}_2(\mathbb{C})/\,\mathrm{SU}(2) &\xrightarrow{\sim} \mathbf{H}^3; \end{aligned}$$

so the fibers of the continuous projection map

$$\mathcal{O}^1\backslash B^1_{\mathbb{R}} \to \Gamma\backslash\mathcal{H} = Y(\Gamma)$$

are given by

$$(\mathbb{H}^1)^{r-t} \times \mathrm{SO}_2(\mathbb{R})^t \times \mathrm{SU}(2)^c$$

and therefore compact, and the claim follows.

Now from (38.2.2), we have

$$\mathcal{O} \hookrightarrow B_{\mathbb{R}} \simeq \mathrm{M}_2(\mathbb{R})^t \times \mathbb{H}^{r-t} \times \mathrm{M}_2(\mathbb{C})^c;$$

choosing an $\mathbb{R}$-basis for $B_{\mathbb{R}}$, we have $B_{\mathbb{R}} \simeq \mathbb{R}^{4n}$ and under the standard metric, this gives $O \simeq \mathbb{Z}^{4n}$ the (non-canonical) structure of a Euclidean lattice. Let $X \subseteq B_{\mathbb{R}}$ be a compact, convex, symmetric subset with volume $\operatorname{vol}(X) > 2^{4n} \operatorname{covol}(O)$. (The precise value of this covolume will not figure in the argument, and anyway would depend on the choice of Euclidean structure; all such structures induce the same topology.) Therefore, by Minkowski's convex body theorem (Theorem 17.5.5), there exists nonzero $\alpha \in O \cap X$. Moreover, for all $g = (g_v)_v \in B_{\mathbb{R}}^1$ we have

$$\operatorname{vol}(gX) = \operatorname{vol}(X)$$

since $\prod_v \det(g_v) = 1$, and gX is again compact, convex, and symmetric, so similarly there exists nonzero $\alpha_g \in O \cap gX$.

We will show that the quotient space $O^1\backslash B_{\mathbb{R}}^1$ is sequentially compact. To this end, let g_n be a sequence from $B_{\mathbb{R}}^1$. By the previous paragraph, there exist $\alpha_n \in O$ such that $\alpha_n = g_n x_n$ with $x_n \in X$ nonzero. Since X is compact, we can restrict to a subsequence such that $x_n \to x \in X$ converges.

The reduced norm $\operatorname{nrd}\colon B \to F$ extends by scaling to a continuous function $\operatorname{nrd}\colon B_{\mathbb{R}} \to F_{\mathbb{R}}$. Since $X \subseteq B_{\mathbb{R}}$ is bounded so too is $\operatorname{nrd}(X) \subseteq F_{\mathbb{R}}$ bounded. But

$$\operatorname{nrd}(\alpha_n) = \operatorname{nrd}(g_n)\operatorname{nrd}(x_n) = \operatorname{nrd}(x_n) \in \operatorname{nrd}(X),$$

and the values $\operatorname{nrd}(\alpha_n) \in \operatorname{nrd}(O) \subseteq \mathbb{Z}_F$ lie in a discrete subset, so there are only finitely many possibilities for $\operatorname{nrd}(\alpha_n)$. Moreover, the left ideals $I_n = O\alpha_n$ have

$$\mathsf{N}(I_n) = \operatorname{Nm}_{F|\mathbb{Q}}(\operatorname{nrd}(I_n))^2 = \operatorname{Nm}_{F|\mathbb{Q}}(\operatorname{nrd}(\alpha_n))^2 \tag{38.4.4}$$

bounded, so there are only finitely many possibilities for I_n by Lemma 17.7.26. Thus, by the pigeonhole principle and restricting to a subsequence, we may suppose $I_n = O\alpha_n = O\alpha_1$ and $\operatorname{nrd}(\alpha_n) = \operatorname{nrd}(\alpha_1)$ for all n. Therefore $\alpha_n = \gamma_n\alpha_1$ with $\gamma_n \in O^1$ for all n.

To conclude, we note that since B is a division algebra,

$$\operatorname{nrd}(x_n) = \operatorname{nrd}(\alpha_n) = \operatorname{nrd}(\alpha_1) \neq 0$$

so

$$x_n^{-1} = \overline{x_n}\operatorname{nrd}(x_n)^{-1} = \overline{x_n}\operatorname{nrd}(\alpha_1)^{-1} \to \overline{x}\operatorname{nrd}(\alpha_1)^{-1} = x^{-1}$$

converges. Therefore

$$\gamma_n^{-1} g_n = \gamma_n^{-1}\alpha_n x_n^{-1} = \alpha_1 x_n^{-1} \to \alpha_1 x^{-1}$$

converges as well. Therefore the quotient $O^1\backslash B_{\mathbb{R}}^1$ is sequentially compact, and therefore compact. □

Remark 38.4.5. Main Theorem 38.4.3 was proven by Hey [Hey29, Hilfssatz 4] in her 1929 Ph.D. thesis (see also Remark 29.10.24) in the case where B is a division algebra over $\mathbb{Q}$. In particular, there is no need to suppose that B is central, so it contains the Dirichlet unit theorem as a consequence: see Exercise 38.6.

After treating the decomposition of the adelic coset space, we will also see Hey's theorem as an essentially direct consequence of Fujisaki's lemma (Main Theorem 27.6.14): see Theorem 38.7.21.

Theorem 38.4.6. *The group* Γ *is finitely generated.*

Proof. Let $x_0 \in X(\Gamma)$ satisfy $\mathrm{Stab}_\Gamma(x_0) = \{1\}$. Then by Theorem 37.4.18, the Dirichlet domain ⌑ $=$ ⌑$(\Gamma; x_0)$ is a locally finite fundamental set for Γ. By Main Theorem 38.4.3, $X(\Gamma)$ is compact, so too is ⌑. Then since ⌑ is locally finite, we conclude that γ⌑ $\cap$ ⌑ $\neq \emptyset$ for only finitely many $\gamma \in \Gamma$. But then by Theorem 37.4.2, the group Γ is generated by such elements and Γ is finitely generated. □

Remark 38.4.7. In fact, Γ is finitely presented, and an argument similar to Proposition 37.3.14 shows that orbits of nonempty subsets

⌑ $\cap\, \gamma$⌑ $\cap\, \gamma'$⌑

with $\gamma \neq \gamma'$ provide a finite generating set of relations (generalizing the vertex cycle relations): see e.g. Raghunathan [Rag72, Theorem 6.15].

38.5 ∗ Arithmetic groups, more generally

In this section, we briefly discuss more general definitions of arithmetic group and show that they reduce to the "working" one given above (Definition 38.3.4). As above, let F be a number field.

38.5.1. Let $\mathsf{G} \leq \mathrm{GL}_{n,F}$ be a **linear algebraic group**, a subgroup variety of $\mathrm{GL}_{n,F}$ defined by polynomial equations in the entries and the inverse of the determinant, with coefficients in F. Equivalently, G is an affine variety over F equipped with identity, multiplication, and inversion morphisms giving it the structure of a group variety.)

We say a subgroup

$$\Gamma \leq \mathsf{G}(F_\infty) = \prod_{v \mid \infty} \mathsf{G}(F_v) \leq \prod_{v\mid\infty} \mathrm{GL}_n(F_v)$$

is **arithmetic** (as a subgroup of $\mathsf{G}(F_\infty)$) if it is commensurable with $\mathsf{G}(\mathbb{Z}_F)$. This notion of arithmetic group was developed significantly by Borel [Bor62, Bor69].

38.5.2. Let B be a quaternion algebra over F. Then there is an embedding $\rho\colon B \hookrightarrow \mathrm{M}_4(F)$ as in Exercise 2.11 by the regular representation. Thus $B^\times \leq \mathrm{GL}_4(F)$ and the image is described by explicit polynomial equations (Exercise 38.5). Therefore there exists a linear algebraic group $\mathsf{G} \leq \mathrm{GL}_{4,F}$ such that $\mathsf{G}(F) \simeq B^\times$.

Similar statements hold for $B^1 \leq \mathrm{SL}_4(F)$.

Lemma 38.5.3. *A group* Γ *commensurable with* $\Gamma^1(\mathcal{O})$ *for a quaternion algebra* B *and order* $\mathcal{O} \subseteq B$ *is arithmetic in the sense of* 38.5.1.

Proof. Applying 38.5.2, we have $\rho\colon B \hookrightarrow \mathrm{M}_4(F)$ realizing $B^1 \simeq \mathsf{G}(F) \leq \mathrm{GL}_4(F)$ as a linear algebraic group by appropriate polynomial equations. Under this embedding, $\mathcal{O} := \rho(B) \cap \mathrm{M}_4(\mathbb{Z}_F)$ is a $\mathbb{Z}_F$-order, and thus $\mathcal{O}^1 \simeq \mathsf{G}(\mathbb{Z}_F)$ and $\prod_v \mathsf{G}(F_v) \simeq \prod_v B_v^1$, as required in the definition. □

In view of Definition 38.3.4, we consider now the converse: when does the more general definition give rise to discrete subgroups of two-by-two matrices?

38.5.4. Let G be a linear algebraic group over F, and suppose that $\mathsf{G}_{\overline{F}} \simeq \mathrm{GL}_{2,\overline{F}}$, so there is a chance to obtain discrete groups of symmetries like the ones considered above. (For a complete treatment, we should consider SL_2 as well as the group PGL_2, but the arguments are similar.) We say then that G is an F-**form** of the algebraic group GL_2.

Lemma 38.5.5. *Let G be an F-form of GL_2. Then there exists a quaternion algebra B over F, unique up to F-algebra isomorphism, such that $\mathsf{G}(F) \simeq B^\times$.*

Proof. This is a basic result in non-abelian Galois cohomology, and it would take us too far afield to prove it here: see e.g. Milne [Milne2017, Theorem 20.3.5] and more generally Serre [Ser79, Chapter X]. ∎

Lemma 38.5.5 explains that more general notions of arithmetic groups do not create anything new beyond our quaternionic definition.

Remark 38.5.6. In this context, there is a criterion for compactness, generalizing Main Theorem 38.4.3 (conjectured by Godement): A discrete subgroup of $\mathsf{G}(\mathbb{R})$ is cocompact if and only if the reductive part of the connected component of G is anisotropic over F. If G is semisimple, then cocompactness is equivalent to asking that every element of $\mathsf{G}(F)$ is semisimple. This criterion was proven by Borel–Harish-Chandra [BHC62] and Mostow–Tamagawa [MT62]; Godement [God62] (with Weil) extended the method of Mostow–Tamagawa and simplified the proof by working directly on adele groups. See also Platonov–Rapinchuk [PR94, §4.5].

38.6 ∗ Modular curves, seen idelically

We have already seen how idelic methods can be both a conceptual and a computational simplification. The quaternion groups defined above naturally also fit into this perspective, and we describe this in the final two sections. As motivation, we begin in this section by reconsidering the classical modular curves from an idelic point of view.

38.6.1. Recall that the adeles of $\mathbb{Q}$ decompose as

$$\underline{\mathbb{Q}} = \widehat{\mathbb{Q}} \times \mathbb{R}$$

into finite and infinite parts. Let $B = \mathrm{M}_2(\mathbb{Q})$. Then

$$\underline{B} = \mathrm{M}_2(\underline{\mathbb{Q}}) = \mathrm{M}_2(\widehat{\mathbb{Q}}) \times \mathrm{M}_2(\mathbb{R}) = \widehat{B} \times B_\infty.$$

The order $O = \mathrm{M}_2(\mathbb{Z})$ is maximal in B, and we have the adelic order $\widehat{O} = \mathrm{M}_2(\widehat{\mathbb{Z}}) \subset \mathrm{M}_2(\widehat{\mathbb{Q}})$. (We have seen that $B = \mathrm{M}_2(\mathbb{Q}) \leq \underline{B} = \mathrm{M}_2(\underline{\mathbb{Q}})$ sits discretely and the quotient $B\backslash\underline{B}$ is compact; like the adelic quotient $F\backslash\underline{F}$ itself, this is not very interesting.)

Similarly, we have

$$\underline{B}^\times = \mathrm{GL}_2(\underline{\mathbb{Q}}) = \widehat{B} \times B_\infty^\times = \mathrm{GL}_2(\widehat{\mathbb{Q}}) \times \mathrm{GL}_2(\mathbb{R}). \tag{38.6.2}$$

It was a key consequence of strong approximation—but easy to establish in this case (Lemma 28.2.4)—that

$$B^\times\backslash\widehat{B}^\times/\widehat{\mathcal{O}}^\times = \mathrm{GL}_2(\mathbb{Q})\backslash\mathrm{GL}_2(\widehat{\mathbb{Q}})/\mathrm{GL}_2(\widehat{\mathbb{Z}}) = \{1\} \tag{38.6.3}$$

is trivial: every nonzero right (invertible) fractional $\mathrm{M}_2(\mathbb{Z})$-ideal is principal, or equivalently, every $\widehat{\mathbb{Z}}$-lattice in $\widehat{\mathbb{Q}}^2$ has a basis in $\mathbb{Q}^2$.

As lovely as this is, this description leaves out the real place, and by putting it back we restore archimedean structure.

38.6.4. The projection map

$$\mathrm{GL}_2(\underline{\mathbb{Q}}) = \mathrm{GL}_2(\widehat{\mathbb{Q}}) \times \mathrm{GL}_2(\mathbb{R}) \to \mathrm{GL}_2(\widehat{\mathbb{Q}})$$

yields a continuous projection

$$\mathrm{GL}_2(\mathbb{Q})\backslash\mathrm{GL}_2(\underline{\mathbb{Q}})/\mathrm{GL}_2(\widehat{\mathbb{Z}}) \to \mathrm{GL}_2(\mathbb{Q})\backslash\mathrm{GL}_2(\widehat{\mathbb{Q}})/\mathrm{GL}_2(\widehat{\mathbb{Z}}) = \{1\}$$

by (38.6.3). Therefore, every element of $\mathrm{GL}_2(\underline{\mathbb{Q}})$ is represented in the double coset by an element of the form $(1, \alpha_\infty)$ with $\alpha_\infty \in \mathrm{GL}_2(\mathbb{R})$; and the element α_∞ is well-defined up the action of the group of pairs $(\gamma, \widehat{\mu}) \in \mathrm{GL}_2(\mathbb{Q}) \times \mathrm{GL}_2(\widehat{\mathbb{Z}})$ satisfying

$$\gamma(1, \alpha_\infty)\widehat{\mu} = (\gamma\widehat{\mu}, \gamma\alpha_\infty) = (1, \alpha'_\infty)$$

so $\gamma = \widehat{\mu}^{-1} \in \mathrm{GL}_2(\mathbb{Q}) \cap \mathrm{GL}_2(\widehat{\mathbb{Z}}) = \mathrm{GL}_2(\mathbb{Z})$, acting on the *left*. In other words, we have a bijection

$$\mathrm{GL}_2(\mathbb{Q})\backslash\mathrm{GL}_2(\underline{\mathbb{Q}})/\mathrm{GL}_2(\widehat{\mathbb{Z}}) \leftrightarrow \mathrm{GL}_2(\mathbb{Z})\backslash\mathrm{GL}_2(\mathbb{R}). \tag{38.6.5}$$

38.6.6. At this point, we are no stranger to the quotient $\mathrm{GL}_2(\mathbb{Z})\backslash\mathrm{GL}_2(\mathbb{R})$! We have studied in detail the related quotient $\mathrm{SL}_2(\mathbb{Z})\backslash\mathrm{SL}_2(\mathbb{R})$: by the symmetric space description, we have an isometry

$$\begin{aligned} \mathrm{SL}_2(\mathbb{R})/\mathrm{SO}(2) &\xrightarrow{\sim} \mathbf{H}^2 \\ g\,\mathrm{SO}(2) &\mapsto gi. \end{aligned} \tag{38.6.7}$$

We similarly obtain a bijection

$$\mathrm{GL}_2(\mathbb{R})/\mathbb{R}^\times\mathrm{SO}(2) \xrightarrow{\sim} \mathbf{H}^{2\pm} = \mathbb{C} \smallsetminus \mathbb{R} \tag{38.6.8}$$

(any matrix of negative determinant interchanges the upper and lower half-planes, so we maintain a bijection). Then we can take the quotient on the left by $\mathrm{GL}_2(\mathbb{Z})$ to get an identification

$$Y(1) = \mathrm{PSL}_2(\mathbb{Z})\backslash\mathbf{H}^2 = \mathrm{GL}_2(\mathbb{Z})\backslash\mathrm{GL}_2(\mathbb{R})/\mathbb{R}^\times\mathrm{SO}(2) \tag{38.6.9}$$

from (38.6.7), and we find again the classical modular curve we considered in section 35.1. (In section 40.1, we will see another version of this in that the space $Y(1)$ parametrizes *complex* lattices $\Lambda \subseteq \mathbb{C}$ up to scaling by $\mathbb{C}^\times$.)

Putting (38.6.9) and (38.6.5) together, we have

$$\begin{aligned} Y(1) &\leftrightarrow \mathrm{GL}_2(\mathbb{Z})\backslash \mathbf{H}^{2\pm} \\ &\leftrightarrow \mathrm{GL}_2(\mathbb{Z})\backslash \mathrm{GL}_2(\mathbb{R})/\mathbb{R}^\times \mathrm{SO}(2) \\ &\leftrightarrow (\mathrm{GL}_2(\mathbb{Q})\backslash \mathrm{GL}_2(\underline{\mathbb{Q}})/\mathrm{GL}_2(\widehat{\mathbb{Z}}))/(\mathbb{R}^\times \mathrm{SO}(2)) \\ &= \mathrm{GL}_2(\mathbb{Q})\backslash(\mathrm{GL}_2(\widehat{\mathbb{Q}})/\mathrm{GL}_2(\widehat{\mathbb{Z}}) \times \mathrm{GL}_2(\mathbb{R})/\mathbb{R}^\times \mathrm{SO}(2)) \\ &= B^\times\backslash(\widehat{B}^\times/\widehat{O}^\times \times B_\infty^\times/K_\infty) = B^\times\backslash\underline{B}^\times/\underline{K}; \end{aligned} \tag{38.6.10}$$

in the last line, we introduce the notation $K_\infty := \mathbb{R}^\times \mathrm{SO}(2)$ and $\underline{K} := \widehat{O}^\times \times K_\infty$ to help in grasping this double coset. Chasing down all of the maps, the bijection is obtained by sending an element $z \in \mathbf{H}^{2\pm}$ to the class of $(1, \alpha) \in \mathrm{GL}_2(\widehat{\mathbb{Q}}) \times \mathrm{GL}_2(\mathbb{R})$ where $z = \alpha i$. Note there is a nice symmetry in the expression on the right-hand side of (38.6.10).

Remark 38.6.11. In this way, we just wrapped the classical quotient in ideles. This double cosetification (a beautiful monster of a word) provides a uniform way to describe the orbifold quotients obtained from quaternionic arithmetic groups more generally: in particular, class number issues are made more transparent in the language of double cosets. On the other hand, geometric structures are not always visible in this language, which is why we have treated both approaches in this text.

38.7 ∗ Double cosets

In this section, we give a description of quaternionic orbifolds in terms of idelic double cosets. We retain the notation from 38.2.1, 38.2.6, and 38.3.1. In particular, F is a number field with r real places and c complex places, B is a quaternion algebra over F that is split at t real places with $t + c > 0$, we have an embedding

$$\iota\colon B \to \mathrm{M}_2(\mathbb{R})^t \times \mathrm{M}_2(\mathbb{C})^c.$$

We have defined

$$\mathcal{H} = (\mathbf{H}^2)^t \times (\mathbf{H}^3)^c$$

where we may also write

$$\mathbf{H}^2 \leftrightarrow \mathrm{SL}_2(\mathbb{R})/\mathrm{SO}(2) \quad \text{and} \quad \mathbf{H}^3 \leftrightarrow \mathrm{SL}_2(\mathbb{C})/\mathrm{SU}(2)$$

by the symmetric space decomposition. We similarly define

$$\mathcal{H}^\pm = (\mathbf{H}^{2\pm})^t \times (\mathbf{H}^3)^c$$

where similarly

$$\mathbf{H}^{2\pm} \leftrightarrow \mathrm{GL}_2(\mathbb{R})/\mathbb{R}^\times \mathrm{SO}(2) \quad \text{and} \quad \mathbf{H}^3 \leftrightarrow \mathrm{GL}_2(\mathbb{C})/\mathbb{C}^\times \mathrm{SU}(2). \tag{38.7.1}$$

In this chapter we have been working with groups $\Gamma \leq \mathrm{Isom}^+(\mathcal{H})$ commensurable with $\Gamma^1(O) = \mathrm{P}\iota(O^1)$. To work idelically, we restrict our attention to such groups that can be defined idelically: these are the congruence subgroups.

Definition 38.7.2. Let $\mathfrak{M}$ be a nonzero ideal. Let

$$O(\mathfrak{M}) := \{\alpha \in O : \alpha \in R + \mathfrak{M}O\}.$$

The **principal congruence subgroup** of level $\mathfrak{M} \subseteq R$ is the group

$$\Gamma(\mathfrak{M}) = \Gamma^1(O(\mathfrak{M})) := \mathrm{P}\iota(O(\mathfrak{M})^1) \leq \mathrm{Isom}^+(\mathcal{H})$$

obtained from the units of norm 1 in $O(\mathfrak{M})$.

A group Γ commensurable with $\Gamma^1(O)$ is **congruence** if it contains $\Gamma(\mathfrak{M})$ for some $\mathfrak{M}$.

A congruence subgroup is defined by (finitely many) congruence conditions. Let $\widehat{\Gamma}$ be the closure of Γ with respect to the topology on $\widehat{B}^\times$, where observe that a fundamental system of open neighborhoods of 1 is given by the images of the principal congruence subgroups.

Lemma 38.7.3. *Suppose that Γ is commensurable with $\Gamma^1(O)$. Then $\widehat{\Gamma} \cap B^\times = \Gamma$ if and only if Γ is congruence.*

Proof. The group $\widehat{\Gamma}$ is a closed subgroup of $\widehat{B}^\times$ commensurable with the compact open subgroup $\widehat{\Gamma(1)}$ and so $\widehat{\Gamma}$ is also compact open. By definition of the topology on $\widehat{B}^\times$, the closure of $\overline{\Gamma} := \widehat{\Gamma} \cap B^\times$ is the smallest congruence group containing Γ, and $\overline{\Gamma}$ contains Γ with finite index. So Γ is congruence if and only if $\Gamma = \overline{\Gamma}$. □

From now on, suppose that Γ is a congruence subgroup.

Remark 38.7.4. In general, we may work with the **congruence closure** $\widehat{\Gamma} \cap B^\times \geq \Gamma$ of an arithmetic group Γ.

38.7.5. In view of (38.6.10) and (38.7.1), we let $K_\infty := \prod_{v\mid\infty} K_v$ where

$$K_v := \begin{cases} \mathbb{R}^\times \mathbb{H}^1 = \mathbb{H}^\times, & \text{if } v \text{ is real and ramified in } B; \\ \mathbb{R}^\times\,\mathrm{SO}(2), & \text{if } v \text{ is real and split in } B; \\ \mathbb{C}^\times\,\mathrm{SU}(2), & \text{if } v \text{ is complex.} \end{cases} \tag{38.7.6}$$

The groups K_v are the extension of a maximal compact subgroup by the center. We then let $\underline{K} = \widehat{\Gamma} \times K_\infty$ and consider the double coset space

$$Y_{\mathrm{Sh}}(\Gamma) := B^\times \backslash \underline{B}^\times / \underline{K} \tag{38.7.7}$$

where $B^\times$ acts on $\underline{B}^\times$ by left multiplication (under the diagonal embedding) and $\underline{K}$ acts on $\underline{B}^\times = \widehat{B}^\times \times B_\infty^\times$ by right multiplication.

The expression (38.7.7) is tidy, and generalizes well, but we also want to know what it looks like. Plugging back in (38.7.1), we have $B_\infty^\times/K_\infty = \mathcal{H}^\pm$, so

$$Y_{\mathrm{Sh}}(\Gamma) = B^\times \backslash ((\widehat{B}^\times/\widehat{\Gamma}) \times \mathcal{H}^\pm) \tag{38.7.8}$$

where $B^\times$ acts on $\mathcal{H}^\pm$ via ι and on $\widehat{B}^\times/\widehat{\Gamma}$ by left multiplication. Since there can hopefully no confusion about this action, removing parentheses we will write

$$Y_{\mathrm{Sh}}(\Gamma) = B^\times \backslash (\widehat{B}^\times \times \mathcal{H}^\pm)/\widehat{\Gamma}.$$

Definition 38.7.9. We call $Y_{\mathrm{Sh}}(\Gamma)$ the **quaternionic Shimura orbifold** of **level** Γ.

Definition 38.7.9 explains the subscript $_{\mathrm{Sh}}$. We immediately proceed to justify the name *orbifold*; moreover, we will see that the space is possibly disconnected, and we write it as a union of connected components.

38.7.10. By weak approximation, there exist elements $\alpha \in B^\times$ with $\operatorname{nrd}(\alpha)$ having all possible real signs at the split real places of B. Therefore

$$Y_{\mathrm{Sh}}(\Gamma) = B^\times_{>0}\backslash(\widehat{B}^\times \times \mathcal{H})/\widehat{\Gamma}. \tag{38.7.11}$$

38.7.12. There is a natural (continuous) projection map

$$Y_{\mathrm{Sh}}(\Gamma) \to B^\times_{>0}\backslash\widehat{B}^\times/\widehat{\Gamma}. \tag{38.7.13}$$

Recall that $t + c > 0$, and B is indefinite. Therefore, strong approximation (as in Corollary 28.6.8) implies that there is a bijection

$$\operatorname{nrd}\colon B^\times_{>0}\backslash\widehat{B}^\times/\widehat{\Gamma} \xrightarrow{\sim} F^\times_{>0}\backslash\widehat{F}^\times/\operatorname{nrd}(\widehat{\Gamma}) =: \mathrm{Cl}^+_{G(\Gamma)} R. \tag{38.7.14}$$

Therefore $\mathrm{Cl}^+_{G(\Gamma)} R$ is a (narrow) class group of F associated to the group Γ; as such, it is a finite abelian group that surjects onto the strict class group $\mathrm{Cl}^+ R$.

38.7.15. By 38.7.12, the space $Y_{\mathrm{Sh}}(\Gamma)$ is the disjoint union of finitely many connected components indexed by the group $\mathrm{Cl}^+_{G(\Gamma)} R$. We identify these connected components explicitly as follows.

Let the ideals $\mathfrak{b} \subseteq R$ form a set of representatives for $\mathrm{Cl}^+_{G(\Gamma)} R$, and let $\widehat{\mathfrak{b}} = \mathfrak{b} \otimes_{\mathbb{Z}} \widehat{\mathbb{Z}}$ be their adelification; then for each $\mathfrak{b}$, there exists $\widehat{b} \in \widehat{R}$ generating $\widehat{\mathfrak{b}}$, so

$$\widehat{b}\widehat{R} \cap R = \mathfrak{b}.$$

For simplicity, choose $\mathfrak{b} = R$ and $\widehat{b} = \widehat{1}$ for the representatives of the trivial class.

By surjectivity of the map (38.7.14), for each $\widehat{b}$ there exists $\widehat{\beta} \in \widehat{B}^\times$ such that $\operatorname{nrd}(\widehat{\beta}) = \widehat{b}$. Therefore

$$Y_{\mathrm{Sh}}(\Gamma) = \bigsqcup_{\mathfrak{b}} B^\times_{>0}(\widehat{\beta}\widehat{\Gamma} \times \mathcal{H}). \tag{38.7.16}$$

For each $\mathfrak{b}$, let

$$\Gamma_{\mathfrak{b}} := \widehat{\beta}\widehat{\Gamma}\widehat{\beta}^{-1} \cap B^\times_{>0} \tag{38.7.17}$$

Then we have a natural bijection

$$\begin{aligned} B^\times_{>0}(\widehat{\beta}\widehat{\Gamma} \times \mathcal{H}) &\leftrightarrow \Gamma_{\mathfrak{b}}\backslash\mathcal{H} \\ (\widehat{\beta}\widehat{\Gamma}, z) &\mapsto z. \end{aligned} \tag{38.7.18}$$

Letting

$$Y(\Gamma_{\mathfrak{b}}) := \Gamma_{\mathfrak{b}}\backslash\mathcal{H}$$

we see that each $Y(\Gamma_{\mathfrak{b}})$ is a connected orbifold of dimension $2t + 3c$. We abbreviate $Y(\Gamma) = Y(\Gamma_{(1)})$ for the trivial class. Putting these together, we have

$$Y_{\mathrm{Sh}}(\Gamma) = \bigsqcup_{\mathfrak{b}} Y(\Gamma_{\mathfrak{b}}) \tag{38.7.19}$$

as a disjoint union of connected orbifolds.

Remark 38.7.20. When $c = 0$, i.e., F is totally real, then $Y_{\mathrm{Sh}}(\Gamma)$ can be canonically given the structure of an algebraic variety defined over a number field, by work of Shimura [Shi67] and Deligne [Del71]; in this case we upgrade $Y_{\mathrm{Sh}}(\Gamma)$ to a **quaternionic Shimura variety**. The theory of Shimura varieties is both broad and deep—see Milne [Milne-SV] and the references therein. We will begin the study quaternionic Shimura varieties of dimension 1 in Chapter 43, touching upon the theory of canonical models in section 43.8.

The above description has hopefully provided a more transparent way to understand arithmetic orbifolds. For example, we can prove an idelic version of Hey's theorem (Main Theorem 38.4.3) as follows.

Theorem 38.7.21. *Suppose that B is a division algebra. Then $Y_{\mathrm{Sh}}(\Gamma)$ is compact.*

Proof. We appeal to Fujisaki's lemma (Main Theorem 27.6.14): the quotient $B^\times\backslash \underline{B}^{(1)}$ is cocompact, under the important hypothesis that B is a division algebra. Let K_∞ be as in 38.7.6, and consider the inclusion followed by the projection

$$B^\times\backslash \underline{B}^{(1)} \hookrightarrow B^\times\backslash \underline{B}^\times \to B^\times\backslash \underline{B}^\times/\underline{K} = Y_{\mathrm{Sh}}(\Gamma). \tag{38.7.22}$$

These maps are continuous, and the composition is surjective as $\mathbb{R}^\times \leq K_\infty$ (embedded diagonally) and $\mathbb{R}^\times \underline{B}^{(1)} = \underline{B}^\times$. So the target is compact, proving the statement. □

Exercises

In these exercises, we maintain the notation in this chapter: let F be a number field with r real places and c complex places, degree $n = [F : \mathbb{Q}]$, and ring of integers R, and let O be an R-order in a quaternion algebra B over F.

1. Let $\alpha \in \mathbb{R} \smallsetminus \mathbb{Q}$. Show that $\mathbb{Z}[\alpha]$ is not discrete in $\mathbb{R}$. (Taking e.g. $\alpha = \sqrt{d}$, this gives a reason to worry about discreteness of number fields when we project.)

2. Embed O^1 diagonally in $\mathrm{SL}_2(\mathbb{R})^{r-t} \times \mathrm{SL}_2(\mathbb{C})^c$. Show that a (further) projection to a proper factor is *not* discrete.

3. Let $B = \left(\dfrac{a, b}{F}\right)$, and let v be a split real place of B. Show that $O^1 \hookrightarrow B_v^1 \simeq \mathrm{SL}_2(\mathbb{R})$ if and only if F is totally real and for all nonidentity real places v', we have $v'(a) < 0$ and $v(b) < 0$.

4. In this exercise, we give a direct argument for the discreteness of an arithmetic Fuchsian group. Suppose F is totally real, let v be a split place of B, consider $F \hookrightarrow v(F) \subseteq \mathbb{R}$ as a subfield of $\mathbb{R}$, and suppose that B is ramified at all other (nonidentity) real places. Prove that $O^1 \subseteq \mathrm{SL}_2(\mathbb{R})$ is discrete.

5. Consider the regular representation $\rho\colon B \hookrightarrow \mathrm{M}_4(F)$ (Exercise 2.11). Describe the image explicitly in terms of polynomial equations in matrix entries. Conclude that $B^\times$ and B^1 are also described by polynomial equations.

(a) Suppose not: then there exists a sequence $\alpha_n = t_n + x_n i + y_n j + z_n ij \to 1$ with $t_n, x_n, y_n, z_n \in F$ with bounded denominators. Multiplying through, suppose that all coordinates are integral. Show for n sufficiently large that all of the coordinates are integral and bounded.
(b) Show that for all nonidentity v, the coordinates of $v(\alpha_n)$ are also bounded using compactness.
(c) Finally, prove that there are only a finite number of elements in R that are bounded in each coordinate (all conjugates are bounded). *[Hint: look at the coefficients of a minimal polynomial, and derive a contradiction.]*

6. Let F be a number field and $R = \mathbb{Z}_F$ its ring of integers. In this exercise, we give a proof of Dirichlet's unit theorem using the same method as in the proof of Main Theorem 38.4.3.

(a) Show that $[R^\times : R^1] \leq 2$.
(b) Show (following the proof of Main Theorem 38.4.3) that $R^1 \backslash F_{\mathbb{R}}^1$ is compact.
(c) Under the usual logarithmic embedding $\log : F_{\mathbb{R}}^1 \to (\prod_v \mathbb{R})^0$, conclude that $\log R^1 \backslash \log F_{\mathbb{R}}^1$ is compact, and therefore $\log R^1$ has rank $r + c - 1$ as an abelian group (written additively).
(d) Conclude that $R^\times$ has rank $r + c - 1$ as an abelian group (written multiplicatively).

This chapter is licensed under the terms of the Creative Commons Attribution-NonCommercial 4.0 International License (http://creativecommons.org/licenses/by-nc/4.0/), which permits any noncommercial use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license and indicate if changes were made.

The images or other third party material in this chapter are included in the chapter's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the chapter's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.

Chapter 39

Volume formula

In this chapter, we exhibit a formula for the covolume of a quaternionic group, a formula with many applications.

39.1 ▹ Statement

We saw in (35.1.5) that the hyperbolic area of the quotient $\mathrm{SL}_2(\mathbb{Z})\backslash\mathbf{H}^2$ can be computed directly from the fundamental domain

$$\mathcal{D} = \{z \in \mathbf{H}^2 : |\mathrm{Re}\, z| \leq 1/2 \text{ and } |z| \geq 1\}$$

as

$$\mathrm{area}(\mathrm{SL}_2(\mathbb{Z})\backslash\mathbf{H}^2) = \mathrm{area}(\mathcal{D}) = \int_{\mathcal{D}} \frac{\mathrm{d}x\,\mathrm{d}y}{y^2} = \frac{\pi}{3}. \tag{39.1.1}$$

In fact, given a Fuchsian or Kleinian group Γ, the hyperbolic area or volume of the quotient $\Gamma\backslash\mathcal{H}$ (where $\mathcal{H} = \mathbf{H}^2, \mathbf{H}^3$, respectively) can be computed *without* recourse to a fundamental domain: it is given in terms of the arithmetic invariants of the order and quaternion algebra that give rise to Γ.

To begin, we consider the already interesting case where the quaternion algebra is defined over $\mathbb{Q}$.

Theorem 39.1.2. *Let B be a quaternion algebra over $\mathbb{Q}$ of discriminant D and let $\mathcal{O} \subseteq B$ be a maximal order. Let $\Gamma_0^1(D) \leq \mathrm{PSL}_2(\mathbb{R})$ be the Fuchsian group associated to the group $\mathrm{P}\mathcal{O}^1 = \mathcal{O}^1/\{\pm 1\}$ of units of reduced norm 1.*

Then

$$\mathrm{area}(\Gamma_0^1(D)\backslash\mathbf{H}^2) = \frac{\pi}{3}\varphi(D) \tag{39.1.3}$$

where

$$\varphi(D) := \prod_{p \mid D}(p-1) = D\prod_{p \mid D}\left(1 - \frac{1}{p}\right) \tag{39.1.4}$$

Recall that the discriminant $D \in \mathbb{Z}$ is a squarefree positive integer, so the function φ is just the Euler totient function. Theorem 39.1.2 recovers (39.1.1) with $D = 1$.

© The Author(s) 2021

J. Voight, *Quaternion Algebras*, Graduate Texts in Mathematics 288,
https://doi.org/10.1007/978-3-030-56694-4_39

Example 39.1.5. Recall the case $X_0^1(6)$ from section 38.1. We confirm that the hyperbolic area computed from the fundamental domain agrees with the formula in Theorem 39.1.2:

$$\operatorname{area}(X_0^1(6,1)) = \frac{\pi}{3}\varphi(6)\psi(1) = \frac{2\pi}{3}.$$

The expression (39.1.3) is quite similar to the Eichler mass formula (Theorem 25.3.15). Indeed, the method of proof is the same, involving the zeta function of the order $\mathcal{O}$. For the case of a definite quaternion algebra, when the unit group was finite, we used Theorem 26.2.12 to relate the mass to a residue of the zeta function (see in particular Proposition 26.5.10); for an indefinite quaternion algebra, to carry this out in general would involve a multivariable integral whose evaluation is similar to the proof of Proposition 26.2.18 (the commutative case)—not an appealing prospect. That being said, this type of direct argument with the zeta function was carried out by Shimizu [Shz65, Appendix] over a totally real field F.

We prefer instead to use idelic methods; we already computed the normalized Tamagawa measure $\tau^1(B^1\backslash\underline{B}^1) = 1$ (see Theorem 29.11.3), and this number has everything we need! It is a much simpler computation to relate this to the volume of the hyperbolic quotient: we carry this out in section 39.3.

The following notation will be used throughout.

39.1.6. Let F be number field of degree $n = [F:\mathbb{Q}]$ with r real places and c complex places, so $r + 2c = n$. Let B be a quaternion algebra over F of discriminant $\mathfrak{D}$ that is split at t real places. Suppose that B is indefinite (i.e., $t + c > 0$).

Let $R = \mathbb{Z}_F$ be the ring of integers of F. Let $\mathcal{O} \subseteq B$ be an R-order of reduced discriminant $\mathfrak{N}$ that is *locally norm-maximal*. Let

$$\mathcal{H} := (\mathbf{H}^2)^t \times (\mathbf{H}^3)^c$$

and let

$$\Gamma^1(\mathcal{O}) \le \mathrm{PSL}_2(\mathbb{R})^t \times \mathrm{PSL}_2(\mathbb{C})^c \circlearrowright \mathcal{H}$$

be the discrete group associated to the group $\mathrm{P}\mathcal{O}^1 = \mathcal{O}^1/\{\pm 1\}$ of units of $\mathcal{O}$ of reduced norm 1.

For a prime $\mathfrak{p} \mid \mathfrak{N}$ with $\operatorname{Nm}(\mathfrak{p}) = q$, let $\left(\dfrac{\mathcal{O}}{\mathfrak{p}}\right) \in \{-1,0,1\}$ be the Eichler symbol (Definition 24.3.2), and let

$$\lambda(\mathcal{O},\mathfrak{p}) := \frac{1-\operatorname{Nm}(\mathfrak{p})^{-2}}{1-\left(\dfrac{\mathcal{O}}{\mathfrak{p}}\right)\operatorname{Nm}(\mathfrak{p})^{-1}} = \begin{cases} 1+1/q, & \text{if } (\mathcal{O}\mid p) = 1; \\ 1-1/q, & \text{if } (\mathcal{O}\mid p) = -1; \\ 1-1/q^2, & \text{if } (\mathcal{O}\mid p) = 0. \end{cases} \tag{39.1.7}$$

Main Theorem 39.1.8. *With notation as in 39.1.6, we have*

$$\operatorname{vol}(\Gamma^1(\mathcal{O})\backslash\mathcal{H}) = \frac{2(4\pi)^t}{(4\pi^2)^r(8\pi^2)^c}\zeta_F(2)d_F^{3/2}\operatorname{Nm}(\mathfrak{N})\prod_{\mathfrak{p}\mid\mathfrak{N}}\lambda(\mathcal{O},\mathfrak{p}). \tag{39.1.9}$$

Main Theorem 39.1.8 is proven as Main Theorem 39.3.1. Since $\zeta_F(2) \approx 1$, for algebras B with fixed signature r, s, t, we can roughly estimate

$$\operatorname{vol}(\Gamma^1(\mathcal{O})) \approx d_F^{3/2} \operatorname{Nm}(\operatorname{discrd} \mathcal{O}). \tag{39.1.10}$$

39.1.11. The constant factor in (39.1.9) is written to help in remembering the formula; multiplying out, we have

$$\frac{2(4\pi)^t}{(4\pi^2)^r(8\pi^2)^c} = \frac{1}{2^{2r+3c-2t-1}\pi^{2r+2c-t}}.$$

Note that the right-hand side of (39.1.9) is independent of the choice of order $\mathcal{O}$ in the genus of $\mathcal{O}$, as it depends only on $\widehat{\mathcal{O}}$.

Remark 39.1.12. If $\mathcal{O}$ is not locally norm-maximal, then one corrects the formula by inserting factors $[R_{\mathfrak{p}}^\times : \operatorname{nrd}(\mathcal{O}_{\mathfrak{p}}^\times)]$ as in (39.2.10); since $R_{\mathfrak{p}}^\times \subseteq \mathcal{O}_{\mathfrak{p}}^\times$ we have $R_{\mathfrak{p}}^{\times 2} \subseteq \operatorname{nrd}(\mathcal{O}_{\mathfrak{p}}^\times)$, so these factors are at most 2 for each $\mathfrak{p}$.

An important special case of Main Theorem 39.1.8 is the case where $\mathcal{O}$ is an Eichler order, generalizing Theorem 39.1.2.

Theorem 39.1.13. *Suppose that* $\mathcal{O} = \mathcal{O}_0(\mathfrak{M})$ *is an Eichler order of level* $\mathfrak{M}$ *and* $\operatorname{disc} B = \mathfrak{D}$, *so* $\mathfrak{N} = \mathfrak{D}\mathfrak{M}$. *Write* $\Gamma_0^1(\mathfrak{M}) = \Gamma^1(\mathcal{O})$ *and* $\Gamma^1(1)$ *for the group associated to a maximal order* $\mathcal{O}(1) \supseteq \mathcal{O}_0(\mathfrak{M})$. *Then*

$$\operatorname{vol}(\Gamma_0^1(\mathfrak{M})\backslash\mathcal{H}) = \frac{2(4\pi)^t}{(4\pi^2)^r(8\pi^2)^c}\zeta_F(2)d_F^{3/2}\varphi(\mathfrak{D})\psi(\mathfrak{M})$$

where

$$\begin{aligned}\varphi(\mathfrak{D}) &= \#(\mathbb{Z}_F/\mathfrak{D})^\times = \operatorname{Nm}\mathfrak{D}\prod_{\mathfrak{p}\mid\mathfrak{D}}\left(1 - \frac{1}{\operatorname{Nm}\mathfrak{p}}\right)\\ \psi(\mathfrak{M}) &= [\Gamma^1(1) : \Gamma_0^1(\mathfrak{M})] = \operatorname{Nm}\mathfrak{M}\prod_{\mathfrak{p}^e\parallel\mathfrak{M}}\left(1 + \frac{1}{\operatorname{Nm}\mathfrak{p}}\right).\end{aligned} \tag{39.1.14}$$

We write $\mathfrak{p}^e \parallel \mathfrak{M}$ to mean $\mathfrak{p}^e$ **exactly divides** $\mathfrak{M}$, i.e., $\operatorname{ord}_{\mathfrak{p}}(\mathfrak{M}) = e$, and we take the product over all prime power divisors of $\mathfrak{M}$ in the definition of ψ.

In particular, when $\mathcal{O}$ is a maximal order, then $\mathfrak{M} = R$ and $\psi(\mathfrak{M}) = 1$.

Remark 39.1.15. Theorem 39.1.13 is often attributed to Borel [Bor81], who derived it under the (highly nontrivial!) assumption that $\tau^1(B^1\backslash \underline{B}^1) = 1$.

Example 39.1.16. Let $\Gamma^1 = \operatorname{PSL}_2(\mathbb{Z}[i])$. Then

$$\operatorname{vol}(X^1) = \frac{2}{8\pi^2}4^{3/2}\zeta_F(2) = \frac{2}{\pi^2}\zeta_F(2) = 0.3053\ldots.$$

This agrees with the computation we did with a fundamental domain (see 36.6.8), since for χ the character 36.6.14 of conductor 4,

$$\frac{2}{\pi^2}\zeta_F(2) = \frac{2}{\pi^2}\zeta(2)L(2,\chi) = \frac{1}{3}L(2,\chi).$$

39.2 Volume setup

In this section, we setup a few calculations in preparation for the volume formula.

Let F be number field with discriminant d_F, let B be a quaternion algebra over F of discriminant $\mathfrak{D} = \operatorname{disc} B$, and let $\mathcal{O} \subseteq B$ be an order. The key input to the proof is the following ingredient: from Theorem 29.11.3,

$$\underline{\tau}^1(B^1\backslash\underline{B}^1) = 1. \tag{39.2.1}$$

We will convert the volume (39.2.1) into the desired form by separating out the contribution from the finite places and the infinite (real) ramified places: what remains is the volume of the orbifold we seek, which we then renormalize from the adelic to the standard hyperbolic volume.

We define

$$\underline{\mathcal{O}} := \widehat{\mathcal{O}} \times B_\infty \subseteq \underline{B}.$$

Then

$$\underline{\mathcal{O}}^1 = \{\underline{\gamma} \in \underline{\mathcal{O}}^\times : \operatorname{nrd}(\underline{\gamma}) = 1\} = \underline{\mathcal{O}}^\times \cap \underline{B}^1$$

and

$$\underline{\mathcal{O}}^1 = \widehat{\mathcal{O}}^1 \times B^1_\infty.$$

Now we apply the important assumption: *we suppose that B is indefinite*. The hypothesis that B is indefinite is necessary even to get a nontrivial space $\mathcal{H} = (\mathbf{H}^2)^t \times (\mathbf{H}^3)^c$ for the group to act upon! The case where B is definite was handled in the proof of the Eichler mass formula (Main Theorem 25.3.19).

39.2.2. By strong approximation (Corollary 28.5.12) we have $\underline{B}^1 = B^1\underline{\mathcal{O}}^1$. Therefore, the natural inclusion

$$\mathcal{O}^1\backslash\underline{\mathcal{O}}^1 \hookrightarrow B^1\backslash\underline{B}^1$$

is also surjective, hence an isomorphism. Thus by (39.2.1)

$$\underline{\tau}^1(\mathcal{O}^1\backslash\underline{\mathcal{O}}^1) = \underline{\tau}^1(B^1\backslash\underline{B}^1) = 1. \tag{39.2.3}$$

We have an embedding $\mathcal{O}^1 \hookrightarrow B^1_\infty$, so

$$1 = \underline{\tau}^1(\mathcal{O}^1\backslash\underline{\mathcal{O}}^1) = \widehat{\tau}^1(\widehat{\mathcal{O}}^1)\tau^1_\infty(\mathcal{O}^1\backslash B^1_\infty). \tag{39.2.4}$$

39.2.5. If $\mathcal{O}$ is maximal, then from (29.7.25) we have

$$\widehat{\tau}^1(\widehat{\mathcal{O}}^1) = \prod_{\mathfrak{p}} \tau^1_{\mathfrak{p}}(\mathcal{O}^1_{\mathfrak{p}}) = |d_F|^{-3/2}\zeta_F(2)^{-1} \prod_{\mathfrak{p}\in\operatorname{Ram}(B)} (\operatorname{Nm}\mathfrak{p} - 1)^{-1}$$

so that

$$\widehat{\tau}^1(\widehat{\mathcal{O}}^1)^{-1} = |d_F|^{3/2}\zeta_F(2)\varphi(\mathfrak{D}). \tag{39.2.6}$$

So by (39.2.6), we conclude that

$$\tau^1_\infty(\mathcal{O}^1\backslash B^1_\infty) = d_F^{3/2}\zeta_F(2)\varphi(\mathfrak{D}). \tag{39.2.7}$$

39.2.8. In general, if $\mathcal{O} \subseteq \mathcal{O}'$ with $\mathcal{O}'$ maximal, then

$$\widehat{\tau}^1(\widehat{\mathcal{O}'}^1) = [\widehat{\mathcal{O}'}^1 : \widehat{\mathcal{O}}^1]\widehat{\tau}^1(\widehat{\mathcal{O}}^1)$$

so similarly

$$\tau^1_\infty(\mathcal{O}^1\backslash B^1_\infty) = d_F^{3/2}\zeta_F(2)\varphi(\mathfrak{D})[\widehat{\mathcal{O}'}^1 : \widehat{\mathcal{O}}^1]. \tag{39.2.9}$$

If $\mathcal{O}$ is locally norm-maximal, then further

$$[\widehat{\mathcal{O}'}^1 : \widehat{\mathcal{O}}^1] = [\widehat{\mathcal{O}'}^\times : \widehat{\mathcal{O}}^\times] \tag{39.2.10}$$

and by Lemma 26.6.7, with $\mathfrak{N} = \operatorname{discrd}\mathcal{O}$ we have

$$\varphi(\mathfrak{D})[\widehat{\mathcal{O}'}^\times : \widehat{\mathcal{O}}^\times] = \prod_{\mathfrak{p}\mid\mathfrak{N}}[\mathcal{O}'_\mathfrak{p} : \mathcal{O}_\mathfrak{p}]\lambda(\mathcal{O},\mathfrak{p}) = \operatorname{Nm}(\mathfrak{N})\prod_{\mathfrak{p}\mid\mathfrak{N}}\lambda(\mathcal{O},\mathfrak{p}) \tag{39.2.11}$$

(as in (26.6.8)).

Next, we relate the measures on $\mathrm{SL}_2(\mathbb{R})$ and $\mathrm{SL}_2(\mathbb{C})$ to the corresponding measures on $\mathbf{H}^2$ and $\mathbf{H}^3$.

39.2.12. Recall the symmetric space identification

$$\mathbf{H}^2 \to \mathrm{SL}_2(\mathbb{R})/\operatorname{SO}(2) \tag{39.2.13}$$

The hyperbolic plane $\mathbf{H}^2$ is equipped with the hyperbolic measure μ, the unique measure invariant under the action of $\mathrm{PSL}_2(\mathbb{R})$; the group $\mathrm{SL}_2(\mathbb{R})$ has the Haar measure τ^1, also invariant under the left action of $\mathrm{SL}_2(\mathbb{R})$. Therefore, the identification (39.2.13) relates these two measures up to a constant $v(\mathrm{SO}(2))$ that gives a total measure to SO(2) (normalizing its Haar measure). Ditto for $\mathbf{H}^3$ and SU(2), with a constant $v(\mathrm{SU}(2))$.

Lemma 39.2.14. *We have*

$$v(\mathrm{SO}(2)) = \pi \qquad \textit{and} \qquad v(\mathrm{SU}(2)) = 8\pi^2.$$

Proof. One could compute the relevant constant by doing a (compatible) integral, but we prefer just to refer to an example where both sides are computed. We consider a fundamental domain for the action of $\mathrm{SL}_2(\mathbb{Z}) \circlearrowright \mathrm{SL}_2(\mathbb{R})$ that is invariant under SO(2): for example, we can lift a fundamental domain for $\mathrm{PSL}_2(\mathbb{Z}) \circlearrowright \mathbf{H}^2$ under (39.2.13). The difference between $\mathrm{SL}_2(\mathbb{Z})$ and $\mathrm{PSL}_2(\mathbb{Z}) = \mathrm{SL}_2(\mathbb{Z})/\{\pm 1\}$ is annoyingly relevant here! We have

$$\mathrm{PSL}_2(\mathbb{Z})\backslash\mathbf{H}^2 = \mathrm{PSL}_2(\mathbb{Z})\backslash\operatorname{SL}_2(\mathbb{R})/\operatorname{SO}(2);$$

if we let $\mathrm{SO}(2)_2 \simeq \mathrm{SO}(2)/\{\pm 1\}$ be the rotation group acting by 2θ instead of θ, then we can lift from $\mathrm{PSL}_2(\mathbb{Z})$ to $\mathrm{SL}_2(\mathbb{Z})$ and

$$\mathrm{PSL}_2(\mathbb{Z})\backslash\operatorname{SL}_2(\mathbb{R})/\operatorname{SO}(2) = \mathrm{SL}_2(\mathbb{Z})\backslash\operatorname{SL}_2(\mathbb{R})/\operatorname{SO}(2)_2$$

which under compatible metrics gives

$$v(\mathrm{SO}(2)) = 2v(\mathrm{SO}(2)_2) = 2\frac{\tau^1(\mathrm{SL}_2(\mathbb{Z})\backslash\operatorname{SL}_2(\mathbb{R}))}{\mu(\mathrm{PSL}_2(\mathbb{Z})\backslash\mathbf{H}^2)}. \tag{39.2.15}$$

On the bottom we have $\mu(\mathrm{SL}_2(\mathbb{Z})\backslash\mathbf{H}^2) = \pi/3$ by (35.1.5) and on the top we have $\tau^1(\mathrm{SL}_2(\mathbb{Z})\backslash\mathrm{SL}_2(\mathbb{R})) = \zeta(2) = \pi^2/6$ by (39.2.9). Plugging in, we compute

$$v(\mathrm{SO}(2)) = 2\frac{\pi^2/6}{\pi/3} = \pi. \tag{39.2.16}$$

Similarly,

$$v(\mathrm{SU}(2)) = 2\frac{\tau^1(\mathrm{SL}_2(\mathbb{Z}[i])\backslash\mathrm{SL}_2(\mathbb{C}))}{\mu(\mathrm{PSL}_2(\mathbb{Z}[i])\backslash\mathbf{H}^3)} = 2\frac{4^{3/2}\zeta_{\mathbb{Q}(i)}(2)}{(2/\pi^2)\zeta_{\mathbb{Q}(i)}(2)} = 8\pi^2 \tag{39.2.17}$$

by Example 39.1.16 and again (39.2.9). $\square$

39.3 Volume derivation

We now establish the volume formula (Main Theorem 39.1.8) using the computation of the Tamagawa measure (Theorem 29.11.3, (39.2.1)), following Borel [Bor81, 7.3].

We continue with the notation from 39.1.6, so in particular $n = [F : \mathbb{Q}]$ and F has r real places, c complex places, and B is split at t real places; B is indefinite, so $t + c > 0$; and $\mathcal{O} \subseteq B$ is a locally norm-maximal order. We restate the formula for convenience.

Main Theorem 39.3.1. *We have*

$$\mathrm{vol}(\Gamma^1(\mathcal{O})\backslash\mathcal{H}) = \frac{2(4\pi)^t}{(4\pi^2)^r(8\pi^2)^c}\zeta_F(2)d_F^{3/2}\,\mathrm{Nm}(\mathfrak{N})\prod_{\mathfrak{p}\mid\mathfrak{N}}\lambda(\mathcal{O},\mathfrak{p}). \tag{39.3.2}$$

Proof. To summarize the previous section, we started with $\tau^1(B^1\backslash\underline{B}^1)$ and concluded $\tau^1(\mathcal{O}^1\backslash\underline{\mathcal{O}}^1) = 1$ by strong approximation; we factored this into finite and infinite parts, with the finite part computed in terms of the order and the infinite part. Then

$$\mathcal{O}^1\backslash B^1_\infty = \mathcal{O}^1\backslash\prod_{v\mid\infty}B^1_v \simeq \prod_{v\in\Omega}B^1_v \times \Bigl(\mathcal{O}^1\backslash\prod_{\substack{v\mid\infty\\ v\notin\Omega}}B^1_v\Bigr). \tag{39.3.3}$$

Each term contributes to the volume. For the first product, for each of the $r-t$ places $v \in \Omega$ we have $B^1_v \simeq \mathbb{H}^1$ and we computed in Lemma 29.5.9 that $\tau^1(\mathbb{H}^1) = 4\pi^2$. For the remaining terms, we employ the comparison formula between measures (Lemma 39.2.14), and are plagued by the same factor 2 coming from the fact that $\Gamma^1(\mathcal{O})$ arises from $\mathrm{P}\mathcal{O}^1/\{\pm1\}$. Putting these together, the decomposition (39.3.3) yields a volume

$$\begin{aligned}\tau^1_\infty(\mathcal{O}^1\backslash B^1_\infty) &= (4\pi^2)^{r-t}\pi^t(8\pi^2)^c\frac{1}{2}\,\mathrm{vol}(\Gamma^1\backslash\mathcal{H})\\ &= \frac{(4\pi^2)^r(8\pi^2)^c}{2(4\pi)^t}\,\mathrm{vol}(\Gamma^1\backslash\mathcal{H}).\end{aligned} \tag{39.3.4}$$

From (39.2.9) and (39.3.4) we conclude

$$\begin{aligned}\mathrm{vol}(\Gamma^1\backslash\mathcal{H}) &= \frac{2(4\pi)^t}{(4\pi^2)^r(8\pi^2)^c}\mu(\mathcal{O}^1\backslash B^1_\infty)\\ &= \frac{2(4\pi)^t}{(4\pi^2)^r(8\pi^2)^c}d_F^{3/2}\zeta_F(2)\varphi(\mathfrak{D})[\widehat{\mathcal{O}'}^1 : \widehat{\mathcal{O}}^1].\end{aligned} \tag{39.3.5}$$

Finally, the computation (39.2.11) of the local index completes the proof. □

Remark 39.3.6. A similar proof works for the case where F is a function field or where Γ is an S-arithmetic group, but in both cases still under the hypothesis that B is S-indefinite for an eligible set S (playing the role of the archimedean places above).

Example 39.3.7. Suppose F is totally real, and we take $B = \mathrm{M}_2(F)$ and $\mathcal{O} = \mathrm{M}_2(\mathbb{Z}_F)$. Then $\mathcal{H} = (\mathbf{H}^2)^n$ and

$$\mathrm{vol}(\Gamma^1(\mathcal{O})\backslash\mathcal{H}) = \frac{2\zeta_F(2)(4\pi)^n}{(4\pi^2)^n} d_F^{3/2} = \frac{2\zeta_F(2)}{\pi^n} d_F^{3/2}.$$

39.4 Genus formula

In this section, we take the volume formula (Main Theorem 39.3.1) in the special case of a Fuchsian group and extend it to a formula for the genus of a Shimura curve.

We maintain our notation but now specialize to the case where $c = 0$ and $t = 1$: in particular, F is a totally real field, and B is indefinite. Thus $\Gamma = \Gamma^1(\mathcal{O}) \leq \mathrm{PSL}_2(\mathbb{R})$ is a Fuchsian group.

We suppose that $\mathcal{O}$ is an Eichler order of level $\mathfrak{M}$.

39.4.1. Recalling 37.7, let $(g; e_1, \dots, e_k; \delta)$ be the signature of Γ. Then $Y(\Gamma)$ has genus g; has k elliptic cycles of orders $e_1, \dots, e_k \in \mathbb{Z}_{\geq 2}$, corresponding to cone points on $Y(\Gamma)$ with given order; and has δ parabolic cycles, corresponding to the punctures of $Y(\Gamma)$. We have $\delta = 0$ unless $B = M_2(\mathbb{Q})$, corresponding to the case of classical modular curves.

By Proposition 37.7.4,

$$\mu(Y(\Gamma)) = 2\pi\left((2g-2) + \sum_{i=1}^{k}\left(1 - \frac{1}{e_i}\right) + \delta\right).$$

Rewriting this slightly, for $q \in \mathbb{Z}_{\geq 2}$, let m_q be the number of elliptic cycles of order q. Then

$$\frac{\mu(Y(\Gamma))}{2\pi} = 2g - 2 + \sum_{q \geq 2} m_q \left(1 - \frac{1}{q}\right) + \delta \tag{39.4.2}$$

and the sum is finite.

The numbers m_q are determined by embedding numbers of quadratic orders into the quaternion order $\mathcal{O}$, as studied in chapter 30.

39.4.3. Let $q \in \mathbb{Z}_{\geq 2}$. Suppose that $m_q > 0$, so that $\mathcal{O}^1$ has a maximal finite subgroup $\langle \gamma \rangle \leq \mathcal{O}^1$ of order $2q$. Then the field $K_q = F(\zeta_{2q}) \supset F$ is a quadratic field extension and $K_q \hookrightarrow B$ embeds, where ζ_{2q} is a primitive $2q$th root of unity, and we have *two* optimal embeddings $S = F(\gamma) \cap \mathcal{O} \hookrightarrow \mathcal{O}$ given by γ and $\overline{\gamma}$. Conversely, to every embedding $\phi\colon K_q \hookrightarrow B$, we associate the order $S = \phi(K_q) \cap \mathcal{O}$ and the finite subgroup $S^\times_{\mathrm{tors}} \subset \mathcal{O}^1$.

Thus there is a two-to-one map

$$\{\mathcal{O}^1\text{-conjugacy classes of optimal embeddings } \phi\colon S \hookrightarrow \mathcal{O} \text{ with } S^\times_{\mathrm{tors}} = 2q\}$$
$$\downarrow$$
$$\{\text{Elliptic cycles of } \Gamma \text{ of order } 2q\}.$$

In the notation of 30.3.10, we have shown that

$$m_q = \frac{1}{2} \sum_{\substack{K_q \supset S \supseteq R[\zeta_{2q}] \\ \#S^\times_{\mathrm{tors}} = 2q}} m(S, \mathcal{O}; \mathcal{O}^1). \tag{39.4.4}$$

Our next major ingredient is the theory of selectivity, treated in chapter 31.

39.4.5. We claim that K_q does not satisfy the selectivity condition (OS), defined in 31.1.6. If $\mathcal{O}$ is an Eichler order, we may appeal to Proposition (31.2.1) and condition (a): since F is totally real, K_q is totally complex, and B is split at a real place, condition (a) fails.

Therefore by Main Theorem 31.1.7(a), $\operatorname{Gen}\mathcal{O}$ is not optimally selective. By Corollary 31.1.10, for every R-order $S \subseteq K_q$, we have

$$m(S, \mathcal{O}; \mathcal{O}^\times) = \frac{h(S)}{\#\operatorname{Cl}_\Omega R} m(\widehat{S}, \widehat{\mathcal{O}}; \widehat{\mathcal{O}}^\times) \tag{39.4.6}$$

where we have substituted $\#\operatorname{Cls}\mathcal{O} = \#\operatorname{Cl}_\Omega R$ (by Corollary 28.5.17).

The adelic embedding numbers $m(\widehat{S}, \widehat{\mathcal{O}}; \widehat{\mathcal{O}}^\times)$, a product of (finitely many) local embedding numbers by 30.7.1, are computed in section 30.6.

We will need one lemma relating units to class numbers.

Lemma 39.4.7. *We have* $[R^\times_{>_\Omega 0} : R^{\times 2}] = 2[\operatorname{Cl}_\Omega R : \operatorname{Cl} R]$.

Proof. The index $[R^\times_{>_\Omega 0} : R^{\times 2}]$, which does not depend on S, is related to class numbers as follows. For each real place v, define $\operatorname{sgn}_v\colon F^\times \to \{\pm 1\}$ by the real sign $\operatorname{sgn}_v(a) = \operatorname{sgn}(v(a))$ at v. Let

$$\begin{aligned} \operatorname{sgn}_\Omega\colon F^\times &\to \{\pm 1\}^\Omega \\ a &\mapsto (\operatorname{sgn}_v(a))_v \end{aligned}$$

collect the signs at the places $v \in \Omega$. Then we have an exact sequence

$$1 \to \{\pm 1\}^\Omega / \operatorname{sgn}_\Omega R^\times \to \operatorname{Cl}_\Omega R \to \operatorname{Cl} R \to 1 \tag{39.4.8}$$

where the map on the left is induced by mapping a tuple of signs in $\{\pm 1\}^\Omega$ to the principal ideal generated by any $a \in F^\times$ with the given signs. We have a second (tautological) exact sequence

$$1 \to R^\times_{>_\Omega 0}/R^{\times 2} \to R^\times/R^{\times 2} \xrightarrow{\operatorname{sgn}_\Omega} \{\pm 1\}^\Omega \to \{\pm 1\}^\Omega / \operatorname{sgn}_\Omega R^\times \to 1 \tag{39.4.9}$$

of elementary abelian 2-groups (or $\mathbb{F}_2$-vector spaces). Combining (39.4.9) with (39.4.8), and noting that $[R^\times : R^{\times 2}] = 2^r$ by Dirichlet's unit theorem and $\#\Omega = r - 1$ by hypothesis, we conclude that

$$[R^\times_{>_\Omega 0} : R^{\times 2}] = 2^{r - \#\Omega}[\operatorname{Cl}_\Omega R : \operatorname{Cl} R] = 2[\operatorname{Cl}_\Omega R : \operatorname{Cl} R]. \qquad \square$$

Definition 39.4.10. For a quadratic R-order $S \subseteq K$, the **Hasse unit index** is defined by

$$Q(S) := [\mathrm{Nm}_{K|F}(S^\times) : R^{\times 2}].$$

We have $Q(S) < \infty$ because $S^\times$ and $R^\times$ have the same $\mathbb{Z}$-rank.

Remark 39.4.11. Hasse [Hass52, Sätze 14–29] proved numerous theorems about $Q(\mathbb{Z}_K)$, including that $Q(\mathbb{Z}_K) \leq 2$: see also Washington [Was97, Theorem 4.12].

We are now ready to write in a simplified way the count m_q of elliptic cycles.

Proposition 39.4.12. *We have*

$$m_q = \frac{1}{h(R)} \sum_{\substack{S \subset K_q \\ \#S^\times_{\mathrm{tors}} = 2q}} \frac{h(S)}{Q(S)} m(\widehat{S}, \widehat{O}; \widehat{O}^\times)$$

where $h(R) = \#\operatorname{Pic} R$ *and* $h(S) = \#\operatorname{Pic} S$.

Proof. Beginning with (39.4.4), we have

$$m_q = \frac{1}{2} \sum_{\substack{S \subset K_q \\ \#S^\times_{\mathrm{tors}} = 2q}} m(S, O; O^1). \tag{39.4.13}$$

By Lemma 30.3.14, we have

$$m(S, O; O^1) = m(S, O; O^\times)[\mathrm{nrd}(O^\times) : \mathrm{nrd}(S^\times)]. \tag{39.4.14}$$

Since B is indefinite, by Corollary 31.1.11 we have $\mathrm{nrd}(O^\times) = R^\times_{>_\Omega 0}$. By Lemma 39.4.7, we have $[R^\times_{>_\Omega 0} : R^{\times 2}] = 2[\mathrm{Cl}_\Omega R : \mathrm{Cl}\, R]$. Thus

$$\begin{aligned} [\mathrm{nrd}(O^\times) : \mathrm{nrd}(S^\times)] &= [R^\times_{>_\Omega 0} : R^{\times 2}][R^{\times 2} : \mathrm{nrd}(S^\times)] \\ &= \frac{2[\mathrm{Cl}_\Omega R : \mathrm{Cl}\, R]}{Q(S)}. \end{aligned} \tag{39.4.15}$$

Substituting (39.4.6) and (39.4.15) into (39.4.14), we find

$$\begin{aligned} m(S, O; O^1) &= m(S, O; O^\times)[\mathrm{nrd}(O^\times) : \mathrm{nrd}(S^\times)] \\ &= \frac{h(S)}{Q(S)} m(\widehat{S}, \widehat{O}; \widehat{O}^\times) \frac{[R^\times_{>_\Omega 0} : R^{\times 2}]}{\#\mathrm{Cl}_\Omega R} \\ &= \frac{2}{h(R)} \frac{h(S)}{Q(S)} m(\widehat{S}, \widehat{O}; \widehat{O}^\times). \end{aligned} \tag{39.4.16}$$

Finally, plugging (39.4.16) into (39.4.13) and cancelling a factor 2 gives the result. □

Corollary 39.4.17. *The signature of a Shimura curve depends only on the discriminant* $\mathfrak{D}$ *and level* $\mathfrak{M}$.

Proof. The ambiguity corresponds to a choice of Eichler order of level $\mathfrak{M}$ and choice of split real place; when $F = \mathbb{Q}$, there is no ambiguity in either case. So we may suppose that Γ has no parabolic cycles. Then we simply observe that the formula (Proposition 39.4.12) for the number of elliptic cycles depends only on $\widehat{O}$. □

Example 39.4.18. As a special case of Proposition 39.4.12, suppose that $\mathfrak{N} = \mathfrak{D}\mathfrak{M}$ is coprime to q. Then as in Example 30.7.5, we have

$$m_q = \frac{1}{h(R)} \prod_{\mathfrak{p} \mid \mathfrak{D}} \left(1 - \left(\frac{K_q}{\mathfrak{p}}\right)\right) \prod_{\mathfrak{p} \mid \mathfrak{M}} \left(1 + \left(\frac{K_q}{\mathfrak{p}}\right)\right) \sum_{\substack{S \subset K_q \\ \# S^\times_{\mathrm{tors}} = 2q}} \frac{h(S)}{Q(S)}.$$

We now have the ingredients to give a formula for a Shimura curve.

Theorem 39.4.19. *Let $Y^1(O) = \Gamma^1(O)\backslash \mathbf{H}^2$. Then $Y^1(O)$ is an orbifold with genus g where*

$$2g - 2 = \frac{4}{(4\pi^2)^r} \zeta_F(2) d_F^{3/2} \varphi(\mathfrak{D}) \psi(\mathfrak{M}) - \sum_{q \geq 2} m_q \left(1 - \frac{1}{q}\right) - \delta$$

where m_q are given in Proposition (39.4.12).

Proof. Combine the volume formula (Main Theorem 39.3.1) with (39.4.2). □

The special case where $F = \mathbb{Q}$ is itself important.

Theorem 39.4.20. *Let $D =$ disc $B > 1$ and let $O \subseteq B$ be an Eichler order of level M, so $N = DM =$ discrd O with D squarefree and* $\gcd(D, M) = 1$.

Then $X^1(O) = \Gamma^1(O)\backslash \mathbf{H}^2$ is an orbifold with genus g where

$$2g - 2 = \frac{\varphi(D)\psi(M)}{6} - \frac{m_2}{2} - \frac{2m_3}{3}$$

where the embedding numbers were computed in Example 30.7.7:

$$m_2 = m(\mathbb{Z}[i], O; O^\times) = \begin{cases} \prod_{p \mid D} \left(1 - \left(\frac{-4}{p}\right)\right) \prod_{p \mid M} \left(1 + \left(\frac{-4}{p}\right)\right), & \text{if } 4 \nmid M; \\ 0, & \text{if } 4 \mid M. \end{cases}$$

$$m_3 = m(\mathbb{Z}[\omega], O; O^\times) = \begin{cases} \prod_{p \mid D} \left(1 - \left(\frac{-3}{p}\right)\right) \prod_{p \mid M} \left(1 + \left(\frac{-3}{p}\right)\right), & \text{if } 9 \nmid M; \\ 0, & \text{if } 9 \mid M; \end{cases}$$

Example 39.4.21. Suppose $D = 6$ and $M = 1$, so we are in the setting of Then $m_2 = m_3 = 2$ and

$$2g - 2 = \frac{\phi(6)}{6} - 1 - \frac{4}{3} = \frac{1}{3} = -2$$

so $g = 0$; this confirms that $X^1(O)$ has signature $(0; 2, 2, 3, 3)$ as in 37.9.10.

Exercises

1. Let F be the function field of a curve X over $\mathbb{F}_q$ of genus g. Let B be a quaternion algebra over F.

   (a) Let $v \in \operatorname{Pl} F$ be place that is *split* in B. Let $S = \{v\}$, let $R = R_{(S)}$, and let $\mathcal{O} \subseteq B$ be an R-order. Let $\mathcal{T}$ be the Bruhat–Tits tree associated to $B_v \simeq \mathrm{M}_2(F_v)$. Via the embedding $\iota : B \hookrightarrow B_v$, show that the group $\Gamma^1(\mathcal{O}) = \mathrm{P}\iota(\mathcal{O}^1) \circlearrowright \mathcal{T}$ acts on $\mathcal{T}$ by left multiplication as a discrete group acting properly.

   (b) Continuing as in (a), compute the measure of $\Gamma^1(\mathcal{O})\backslash\mathcal{T}$ using the methods of section 39.3.

2. Generalizing the previous exercise, let F be a global field, let B be a quaternion algebra over F, let S be an eligible set and suppose that B is S-indefinite. Let $R = R_{(S)}$ and let $\mathcal{O} \subseteq B$ be an R-order. Define a symmetric space $\mathcal{H}$ on which $\mathrm{P}\mathcal{O}^1$ acts as a discrete group acting properly, and compute the measure of $\Gamma^1(\mathcal{O})\backslash\mathcal{H}$.

This chapter is licensed under the terms of the Creative Commons Attribution-NonCommercial 4.0 International License (http://creativecommons.org/licenses/by-nc/4.0/), which permits any noncommercial use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license and indicate if changes were made.

The images or other third party material in this chapter are included in the chapter's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the chapter's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.

Part V
Arithmetic geometry

Chapter 40

Classical modular forms

In this chapter, we introduce modular forms on the classical modular group. This chapter will provide motivation as well as important examples for generalizations in this last part of the text.

40.1 ▹ Functions on lattices

In this section, we pursue the interpretation of the quotient $\Gamma\backslash\mathbf{H}^2$ with $\Gamma = \mathrm{PSL}_2(\mathbb{Z})$ as a moduli space of lattices, and we study functions on the quotient. There are a wealth of references for the classical modular forms, including Apostol [Apo90, Chapters 1–2], Diamond–Shurman [DS2005], Miyake [Miy2006, Chapter 4], Lang [Lang95, §1], and Serre [Ser73, Chapter VII]. For this section, see Silverman [Sil2009, Chapter VI] for the complex analytic theory of elliptic curves and the relationship to Eisenstein series.

Recall from 35.3.3 that $Y = \Gamma\backslash\mathbf{H}^2$ parametrizes complex lattices up to homothety, i.e., there is a bijection

$$\begin{aligned} Y = \Gamma\backslash\mathbf{H}^2 &\to \{\Lambda \subset \mathbb{C} \text{ lattice}\}/\sim \\ \Gamma\tau &\mapsto [\mathbb{Z} + \mathbb{Z}\tau]. \end{aligned} \tag{40.1.1}$$

In particular, the set of homothety classes has a natural structure of a Riemann surface, and we seek now to make this explicit. We show that there are natural, holomorphic functions on the set of lattices that allow us to go beyond the bijection 40.1.1 to realize the complex structure on Y explicitly.

Let $\Lambda \subset \mathbb{C}$ be a lattice. To write down complex moduli, we average over Λ in a convergent way, as follows.

Definition 40.1.2. The **Eisenstein series** of weight $k \in \mathbb{Z}_{>2}$ for Λ is

$$G_k(\Lambda) = \sum_{\substack{\lambda\in\Lambda \\ \lambda\neq 0}} \frac{1}{\lambda^k}.$$

If k is odd, then $G_k(\Lambda) = 0$, so let $k \in 2\mathbb{Z}_{\geq 2}$.

© The Author(s) 2021

J. Voight, *Quaternion Algebras*, Graduate Texts in Mathematics 288,
https://doi.org/10.1007/978-3-030-56694-4_40

Lemma 40.1.3. *The series $G_k(\Lambda)$ converges absolutely.*

Proof. Up to homothety (which does not affect convergence), we may suppose $\Lambda = \mathbb{Z} + \mathbb{Z}\tau$, with $\tau \in \mathbf{H}^2$. Then we consider the corresponding absolute sum

$$\sum_{\substack{\lambda \in \mathbb{Z}+\mathbb{Z}\tau \\ \lambda \neq 0}} \frac{1}{|\lambda|^k} = \sum_{\substack{m,n \in \mathbb{Z} \\ (m,n)\neq(0,0)}} \frac{1}{|m+n\tau|^k}. \tag{40.1.4}$$

The number of pairs (m, n) with $r \leq |m\tau + n| < r + 1$ is the number of lattice points in an annulus of area $\pi(r+1)^2 - \pi r^2 = O(r)$, so there are $O(r)$ such points; and thus the series (40.1.4) is majorized by (a constant multiple of) $\sum_{r=1}^{\infty} r^{1-k}$, which is convergent for $k > 2$. □

40.1.5. For $z \in \mathbf{H}^2$ and $k \in 2\mathbb{Z}_{\geq 2}$, define

$$G_k(z) = G_k(\mathbb{Z} + \mathbb{Z}z) = \sum_{\substack{m,n \in \mathbb{Z} \\ (m,n)\neq(0,0)}} \frac{1}{(m+nz)^k}. \tag{40.1.6}$$

Lemma 40.1.7. *$G_k(z)$ is holomorphic for $z \in \mathbf{H}^2$, and*

$$G_k(\gamma z) = (cz+d)^k G_k(z)$$

for all $\gamma = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in \mathrm{PSL}_2(\mathbb{Z})$.

Proof. This is true for $z \in$ ⌑ since then

$$|m+nz|^2 = m^2 + 2mn\operatorname{Re} z + n^2|z|^2 \geq m^2 - mn + n^2 = |m+n\omega|^2$$

thus $|G_k(z)| \leq |G_k(\omega)|$ and so by the Weierstrass M-test, $G_k(z)$ is holomorphic for $z \in$ ⌑: by Morera's theorem, uniform convergence implies holomorphicity. But now for all $\gamma = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in \Gamma$, we claim that

$$G_k(\gamma z) = (cz+d)^k G_k(z) \tag{40.1.8}$$

(and note this does not depend on the choice of sign): indeed,

$$\frac{1}{m + n(\gamma z)} = \frac{cz+d}{(bn+dm) + (an+cm)z} \tag{40.1.9}$$

and the map

$$(n, m) \mapsto (n, m)\begin{pmatrix} a & b \\ c & d \end{pmatrix} = (an+cm, bn+dm)$$

is a permutation of $\mathbb{Z}^2 - \{(0,0)\}$, so by absolute convergence we may rearrange the sum to get

$$\begin{aligned} G_k(\gamma z) &= (cz+d)^k \sum_{\substack{m,n\in\mathbb{Z} \\ (m,n)\neq(0,0)}} \frac{1}{((bn+dm)+(an+cm)z)^k} \\ &= (cz+d)^k \sum_{\substack{m,n\in\mathbb{Z} \\ (m,n)\neq(0,0)}} \frac{1}{(m+nz)^k} = (cz+d)^k G_k(z). \end{aligned} \tag{40.1.10}$$

By transport, since $\Gamma⌑ = \mathbf{H}^2$, we see that $G_k(z)$ is holomorphic on all of $\mathbf{H}^2$. □

40.1.11. In this (somewhat long) paragraph, we connect the theory of Eisenstein series above to the theory of elliptic curves.

Let $\Lambda \subset \mathbb{C}$ be a lattice. We define the **Weierstrass $\wp$-function** (relative to Λ) by

$$\wp(z) = \wp(z;\Lambda) = \frac{1}{z^2} + \sum_{\substack{\lambda\in\Lambda \\ \lambda\neq 0}} \left(\frac{1}{(z-\lambda)^2} - \frac{1}{\lambda^2}\right). \tag{40.1.12}$$

We have

$$\left|\frac{1}{(z-\lambda)^2} - \frac{1}{\lambda^2}\right| \leq \frac{|z|(2|\lambda|+|z|)}{|\lambda|^2(|\lambda|-|z|)^2} = O\left(\frac{1}{|\lambda|^3}\right) \tag{40.1.13}$$

so as above we see that $\wp(z)$ is absolutely convergent for all $z \in \mathbb{C} \smallsetminus \Lambda$ and uniformly convergent on compact subsets, and so defines a holomorphic function on $\mathbb{C} \smallsetminus \Lambda$. Since

$$\frac{1}{(z-\lambda)^2} - \frac{1}{\lambda^2} = \sum_{n=1}^{\infty} (n+1)\frac{z^n}{\lambda^{n+2}} \tag{40.1.14}$$

by differentiating the geometric series, we find

$$\wp(z) = \frac{1}{z^2} + \sum_{k=3}^{\infty} (k-1)G_k(\Lambda)z^k = \frac{1}{z^2} + 3G_4(\Lambda)z^4 + 5G_6(\Lambda)z^6 + \dots. \tag{40.1.15}$$

Differentiating with respect to z and squaring, we find that

$$\left(\frac{\mathrm{d}\wp}{\mathrm{d}z}(z)\right)^2 = \frac{4}{z^6} - \frac{24G_4(\Lambda)}{z^2} - 80G_6(\Lambda) + \dots. \tag{40.1.16}$$

Expanding out the first few terms, we find that

$$f(z) = \left(\frac{\mathrm{d}\wp}{\mathrm{d}z}(z)\right)^2 - 4\wp(z)^3 + 60G_4(\Lambda)\wp(z) + 140G_6(\Lambda) = O(z^2)$$

is holomorphic at $z=0$ and satisfies $f(z+\lambda) = f(z)$ for all $\lambda \in \Lambda$. By periodicity, $f(z)$ takes its maximum in a fundamental parallelogram for Λ; then by Liouville's theorem,

f is bounded on $\mathbb{C}$ so constant. Since $f(0) = 0$, we conclude that $f(z)$ is identically zero.

Following convention, write

$$g_4 = g_4(\Lambda) = 60G_4(\Lambda) \quad \text{and} \quad g_6 = g_6(\Lambda) = 140G_6(\Lambda)$$

and

$$x(z) = \wp(z;\Lambda) \quad \text{and} \quad y(z) = \frac{\mathrm{d}\wp}{\mathrm{d}z}(z;\Lambda).$$

Then the image of the map

$$\begin{aligned} \mathbb{C}/\Lambda &\to \mathbb{P}^2(\mathbb{C}) \\ z &\mapsto (x(z) : y(z) : 1) \end{aligned} \tag{40.1.17}$$

is cut out by the affine equation

$$y^2 = 4x^3 - g_4 x - g_6;$$

the map (40.1.17) is an isomorphism of Riemann surfaces. Looking ahead to Definition 42.1.1, this map exhibits $\mathbb{C}/\Lambda$ as an elliptic curve over $\mathbb{C}$.

To produce holomorphic functions that are well-defined on the quotient $\Gamma\backslash\mathbf{H}^2$, we can take ratios of Eisenstein series; soon we will exhibit a map

$$j : \mathbf{H}^2 \to \mathbb{C} \tag{40.1.18}$$

obtained in this way that defines a bijective holomorphic map $\Gamma\backslash\mathbf{H}^2 \xrightarrow{\sim} \mathbb{C}$ (Theorem 40.3.8).

40.1.19. Eisenstein series can also be thought of as weighted averages over the (cosets of the) group $\mathrm{PSL}_2(\mathbb{Z})$ as follows.

Let $\Gamma_\infty \le \Gamma = \mathrm{PSL}_2(\mathbb{Z})$ be the stabilizer of ∞; then Γ_∞ is the infinite cyclic group generated by $T = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}$. We consider the cosets $\Gamma_\infty\backslash\Gamma$: for $t \in \mathbb{Z}$ and $\gamma = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$ we have $T^t\gamma = \begin{pmatrix} a+tc & b+td \\ c & d \end{pmatrix}$ with the same bottom row. Thus the function $(cz+d)^2$ is well-defined on the coset $\Gamma_\infty\gamma$. Thus we can form the sum

$$E_k(z) = \sum_{\Gamma_\infty\gamma \in \Gamma_\infty\backslash\Gamma} (cz+d)^{-k} = \frac{1}{2} \sum_{\substack{c,d\in\mathbb{Z} \\ \gcd(c,d)=1}} \frac{1}{(cz+d)^k}, \tag{40.1.20}$$

the factor 2 coming from the choice of sign in $\mathrm{PSL}_2(\mathbb{Z})$. Since every nonzero $(m,n) \in \mathbb{Z}^2$ can be written $(m,n) = r(c,d)$ with $r = \gcd(m,n) > 0$ and $\gcd(c,d) = 1$, we find that

$$G_k(z) = \zeta(k)E_k(z).$$

40.2 ▹ Eisenstein series as modular forms

In the previous section, we saw that natural sums (Eisenstein series) defined functions on $\mathbf{H}^2$ that transformed with respect to $\mathrm{PSL}_2(\mathbb{Z})$ with a natural invariance. In this section, we pursue this more systematically.

Definition 40.2.1. Let $k \in 2\mathbb{Z}$ and let $\Gamma \leq \mathrm{PSL}_2(\mathbb{R})$ be a Fuchsian group. A map $f\colon \mathbf{H}^2 \to \mathbb{C} \cup \{\infty\}$ is**weight k-invariant** under Γ if

$$f(\gamma z) = (cz + d)^k f(z) \quad \text{for all } \gamma = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in \Gamma. \tag{40.2.2}$$

40.2.3. If f is weight k invariant and f' is weight k' invariant, then ff' is weight $k + k'$ invariant, and if $k' = k$ then $f + f'$ is weight k invariant. Therefore, the set of weight k-invariant functions has the structure of a $\mathbb{C}$-vector space.

40.2.4. Weight k invariance under Γ can be checked on a set of generators for Γ lifted to $\mathrm{SL}_2(\mathbb{Z})$, as follows. For $\gamma = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in \mathrm{SL}_2(\mathbb{R})$, define $\jmath(\gamma; z) = cz + d$. Then (40.2.2) can be rewritten $f(\gamma z) = \jmath(\gamma; z)^k f(z)$.

For $\gamma' \in \Gamma$, we compute that $\jmath$ satisfies the **cocycle relation**

$$\jmath(\gamma\gamma'; z) = \jmath(\gamma; \gamma' z)\jmath(\gamma'; z) \tag{40.2.5}$$

because if $\gamma' = \begin{pmatrix} a' & b' \\ c' & d' \end{pmatrix}$, then

$$(a'c + c'd)z + b(b'c + dd') = \left(c\left(\frac{a'z + b'}{c'z + d'}\right) + d\right)(c'z + d'). \tag{40.2.6}$$

Therefore, if f is a map with $f(\gamma z) = \jmath(\gamma; z)^k f(z)$ and $f(\gamma' z) = \jmath(\gamma'; z)^k f(z)$, then

$$\begin{aligned} f(\gamma(\gamma' z)) &= \jmath(\gamma; \gamma' z)^k f(\gamma' z) = \jmath(\gamma; \gamma' z)^k \jmath(\gamma'; z)^k f(z) \\ &= \jmath(\gamma\gamma'; z)^k f(z). \end{aligned} \tag{40.2.7}$$

Since $\mathrm{PSL}_2(\mathbb{Z})$ is generated by S, T, it follows from (40.2.7) that a map f is weight k invariant for $\mathrm{PSL}_2(\mathbb{Z})$ if and only if both equalities

$$\begin{aligned} f(z + 1) &= f(z) \\ f(-1/z) &= z^k f(z) \end{aligned} \tag{40.2.8}$$

hold for all $z \in \mathbf{H}^2$.

40.2.9. Since

$$\frac{\mathrm{d}(\gamma z)}{\mathrm{d}z} = \frac{1}{(cz + d)^2} \tag{40.2.10}$$

the weight k invariance (40.2.2) of a map f can be rewritten

$$f(\gamma z)\,\mathrm{d}(\gamma z)^{\otimes k/2} = f(z)\,\mathrm{d}z^{\otimes k/2} \tag{40.2.11}$$

so equivalently, the differential $f(z)\,\mathrm{d}z^{\otimes k/2}$ is (straight up) *invariant* under Γ.

40.2.12. Let $f\colon \mathbf{H}^2 \to \mathbb{C}$ be a meromorphic map that is weight k invariant under a Fuchsian group $\Gamma \ni \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}$. Then $f(z+1) = f(z)$. If f admits a Fourier series expansion in $q = \exp(2\pi i z)$ of the form

$$f(z) = \sum_{n=-\infty}^{\infty} a_n q^n \in \mathbb{C}((q)) \tag{40.2.13}$$

with $a_n \in \mathbb{C}$ and $a_n = 0$ for all but finitely many $n < 0$, then we say that f is **meromorphic** at ∞; if further $a_n = 0$ for $n < 0$, we say f is **holomorphic at** ∞.

More generally, let $\Gamma \leq \mathrm{PSL}_2(\mathbb{Z})$ be a subgroup of finite index. For $\gamma \in \mathrm{PSL}_2(\mathbb{Z})$, we define

$$f[\gamma]_k(z) := j(\gamma; z)^{-k} f(\gamma z). \tag{40.2.14}$$

Then $f[\gamma]_k(z)$ is weight k invariant under the group $\gamma^{-1}\Gamma\gamma$. We say that f is **meromorphic at the cusps** if for every $\gamma \in \mathrm{PSL}_2(\mathbb{Z})$, the function $f[\gamma]_k$ is meromorphic at ∞, in the above sense. Since f is weight k invariant, to check if f is meromorphic at the cusps, it suffices to take representatives of the finite set of cosets $\Gamma\backslash \mathrm{PSL}_2(\mathbb{Z})$. (The name *cusp* comes from the geometric description at ∞ coming from the parabolic stabilizer group, recalling Definition 33.4.5.)

Finally, we say that f is **holomorphic at the cusps** if $f[\gamma]_k(z)$ is holomorphic at ∞ for all $\gamma \in \mathrm{PSL}_2(\mathbb{Z})$, and **vanishes at the cusps** if $f[\gamma]_k(\infty) = 0$ for all γ.

Definition 40.2.15. Let $k \in 2\mathbb{Z}$ and let $\Gamma \leq \mathrm{PSL}_2(\mathbb{Z})$ be a subgroup of finite index. A **meromorphic modular form of weight** k is a meromorphic map $f\colon \mathbf{H}^2 \to \mathbb{C}$ that is weight k invariant under Γ and meromorphic at the cusps. A **meromorphic modular function** is a meromorphic modular form of weight 0.

A **(holomorphic) modular form of weight** k is a holomorphic map $f\colon \mathbf{H}^2 \to \mathbb{C}$ that is weight k invariant under Γ and holomorphic at the cusps. A **cusp form of weight** k is a holomorphic modular form of weight k that vanishes at the cusps.

Let $M_k(\Gamma)$ be the $\mathbb{C}$-vector space of modular forms of weight k for Γ, and let $S_k(\Gamma) \subseteq M_k(\Gamma)$ be the subspace of cusp forms.

Lemma 40.2.16. *The Eisenstein series $G_k(z)$ is a holomorphic modular form of weight $k \in 2\mathbb{Z}_{\geq 2}$ for $\mathrm{PSL}_2(\mathbb{Z})$ with Fourier expansion*

$$G_k(z) = 2\zeta(k) + 2\frac{(2\pi i)^k}{(k-1)!}\sum_{n=1}^{\infty} \sigma_{k-1}(n) q^n \tag{40.2.17}$$

where

$$\zeta(k) = \sum_{n=1}^{\infty} \frac{1}{n^k}$$

and

$$\sigma_{k-1}(n) = \sum_{\substack{d \mid n \\ d>0}} d^{k-1}.$$

Proof. We start with the formula

$$\pi \cot(\pi z) = \sum_{m=-\infty}^{\infty} \frac{1}{z+m} = \lim_{M\to\infty} \sum_{m=-M}^{M} \frac{1}{z+m} \tag{40.2.18}$$

(Exercise 40.2); with $q = \exp(2\pi i z)$,

$$\cot(\pi z) = \frac{\cos(\pi z)}{\sin(\pi z)} = i\frac{q+1}{q-1} = i\left(1 + \frac{2}{q-1}\right)$$

and so we obtain the Fourier expansion

$$\pi \cot(\pi z) = \pi i - 2\pi i \sum_{n=0}^{\infty} q^n. \tag{40.2.19}$$

Equating (40.2.18)–(40.2.19) and differentiating $k-1$ times, we find that

$$\sum_{m=-\infty}^{\infty} \frac{1}{(m+z)^k} = \frac{1}{(k-1)!}(2\pi i)^k \sum_{n=1}^{\infty} n^{k-1} q^n \tag{40.2.20}$$

(since k is even). Thus

$$G_k(z) = \sum_{(m,n)\neq(0,0)} \frac{1}{(m+nz)^k} = 2\zeta(k) + 2\sum_{n=1}^{\infty}\sum_{m=-\infty}^{\infty} \frac{1}{(m+nz)^k}$$

so replacing $n \leftarrow a$ and then substituting $z \leftarrow nz$ in (40.2.20), summing over n we obtain

$$\begin{aligned} G_k(z) &= 2\zeta(k) + 2\frac{(2\pi i)^k}{(k-1)!} \sum_{n=1}^{\infty}\sum_{a=1}^{\infty} a^{k-1} q^{an} \\ &= 2\zeta(k) + 2\frac{(2\pi i)^k}{(k-1)!} \sum_{n=1}^{\infty}\Bigl(\sum_{d\mid n} d^{k-1}\Bigr) q^n \end{aligned} \tag{40.2.21}$$

grouping together terms in the second step. The fact that G_k is holomorphic at ∞ then follows by definition. □

40.2.22. We accordingly define the **normalized Eisenstein series** by

$$E_k(z) = \frac{1}{2\zeta(k)} G_k(z)$$

(see also 40.1.19). We have

$$E_k(z) = 1 - \frac{2k}{B_k} \sum_{n=1}^{\infty} \sigma_{k-1}(n) q^n \tag{40.2.23}$$

where $B_k \in \mathbb{Q}^\times$ are the Taylor coefficients of

$$\frac{x}{e^x - 1} = \sum_{k=0}^{\infty} B_k \frac{x^k}{k!} = 1 - \frac{1}{2}x + \frac{1}{6}\frac{x^2}{2!} - \frac{1}{30}\frac{x^4}{4!} + \frac{1}{42}\frac{x^6}{6!} + \dots$$

(Exercise 40.3): the numbers $B_k \in \mathbb{Q}$ (with $B_k \neq 0$ for $k \in 2\mathbb{Z}_{\geq 0}$) are **Bernoulli numbers**. Expanding, we find

$$\begin{aligned} E_4(z) &= 1 + 240q + 2160q^2 + 6720q^3 + 17520q^4 + \dots \\ E_6(z) &= 1 - 504q - 16632q^2 - 122976q^3 - 532728q^4 - \dots . \end{aligned}$$

Remark 40.2.24. The notion of Eisenstein series extends in a natural way to the Bianchi groups $\mathrm{PSL}_2(\mathbb{Z}_F)$ where F is an imaginary quadratic field: see Elstrodt–Grunewald–Mennicke [EGM98, Chapter 3].

40.3 ▹ Classical modular forms

In this section, we study modular forms for $\mathrm{PSL}_2(\mathbb{Z})$; to ease notation, we abbreviate $\Gamma := \mathrm{PSL}_2(\mathbb{Z})$.

40.3.1. Let f be a meromorphic modular form of weight k for $\Gamma = \mathrm{PSL}_2(\mathbb{Z})$. If $k = 0$, so f is a modular function, then f descends to a meromorphic function on $Y := \Gamma\backslash\mathbf{H}^2$. Although this is not true for (nonzero) forms of weight $k \neq 0$, the *order* of zero or pole $\mathrm{ord}_z(f)$ is well defined on the orbit Γz by weight k invariance (40.2.2). With the Fourier expansion (40.2.13), we define

$$\mathrm{ord}_\infty(f) := \mathrm{ord}_q\Bigl(\sum_n a_n q^n\Bigr) = \min(\{n : a_n \neq 0\}).$$

The form f has only finitely many zeros or poles in Y, i.e., only finitely many Γ-orbits of zeros or poles: since f is meromorphic at ∞, there exists $\epsilon > 0$ such that f has no zero or pole with $0 < |q| < \epsilon$, so with

$$\mathrm{Im}\, z > M = \frac{\log(1/\epsilon)}{2\pi};$$

but the part of ⌑ with $\mathrm{Im}\, z \leq M$ is compact, and since f is meromorphic in $\mathbf{H}^2$, it has only finitely many zeros or poles in this part as well.

40.3.2. In a similar way, the order of the stabilizer $e_z := \#\,\mathrm{Stab}_\Gamma(z)$ is well defined on the orbit Γz, since points in the same orbit have conjugate stabilizers. By 35.1.14,

$$e_z = \begin{cases} 3, & \text{if } \Gamma z = \Gamma\omega; \\ 2, & \text{if } \Gamma z = \Gamma i; \\ 1, & \text{otherwise.} \end{cases} \tag{40.3.3}$$

Proposition 40.3.4. *Let $f \colon \mathbf{H}^2 \to \mathbb{C}$ be a meromorphic modular form of weight k for $\Gamma = \mathrm{PSL}_2(\mathbb{Z})$, not identically zero. Then*

$$\mathrm{ord}_\infty(f) + \sum_{\Gamma z \in \Gamma\backslash\mathbf{H}^2} \frac{1}{e_z}\,\mathrm{ord}_z(f) = \frac{k}{12} \tag{40.3.5}$$

where $e_z := \#\,\mathrm{Stab}_\Gamma(z)$.

The sum (40.3.5) has only finitely many terms, by 40.3.1, and the stabilizers are given in 40.3.2.

Proof. See Serre [Ser73, §3, Theorem 3]: the proof consists of performing a contour integration $\dfrac{1}{2\pi i}\dfrac{\mathrm{d}f}{f}$ on the boundary of ⌑. Alternatively, this statement can be seen as a manifestation of the Riemann–Roch theorem: see Diamond–Shurman [DS2005, §3.5]. □

40.3.6. We have $E_4(STz) = (z+1)^4 E_4(z)$, so since $(ST)(\omega) = \omega$,

$$E_4(\omega) = (\omega+1)^4 E_4(\omega) = \omega^2 G_4(\omega)$$

so $E_4(\omega) = 0$. Since E_4 is holomorphic in $\mathbf{H}^2$, we have $\operatorname{ord}_z(E_4) \in \mathbb{Z}_{\geq 0}$ for all $z \in \mathbf{H}^2$, and thus by Proposition 40.3.4, we must have that $E_4(z)$ has no other zeros in ⌑. Similarly,

$$E_6(i) = E_6(Si) = i^6 E_6(i) = -E_6(i)$$

so $E_6(i) = 0$, and $E_6(z)$ has no other zeros.

For the same reason, the function

$$\Delta(z) = \frac{E_4(z)^3 - E_6(z)^2}{1728} = q - 24q^2 + 252q^3 - 1472q^4 + \dots \tag{40.3.7}$$

is a modular form of weight 12 with no zeros in $\mathbf{H}^2$ with $\operatorname{ord}_\infty(\Delta) = 1$.

We give two applications of Proposition 40.3.4. First, we obtain the identification promised in (40.1.18).

Theorem 40.3.8. *The function*

$$j(z) = \frac{E_4(z)^3}{\Delta(z)} = \frac{1}{q} + 744 + 196884q + 21493760q^2 + \dots \tag{40.3.9}$$

is a meromorphic modular function for $\mathrm{PSL}_2(\mathbb{Z})$*, holomorphic in* $\mathbf{H}^2$*, defining a bijection*

$$Y = \Gamma\backslash\mathbf{H}^2 \to \mathbb{C}.$$

Proof. The function j is weight 0 invariant under Γ as the ratio of two forms that are weight 12 invariant. Since E_4 is holomorphic in $\mathbf{H}^2$, and Δ is holomorphic and has no zeros in $\mathbf{H}^2$, the ratio is holomorphic in $\mathbf{H}^2$ and $j(z)$ has a simple pole at $z = \infty$, corresponding to a simple zero of Δ at $z = \infty$. From 40.3.6, we have $j(i) = 1728$, and $j(z) - 1728$ has a double zero at $z = i$, and $j(z)$ has a triple zero at $z = \omega$.

To conclude that j is bijective, we show that $j(z) - c$ has a unique zero $\Gamma z \in Y$. If $c \neq 0, 1728$, this follows immediately from Proposition 40.3.4; if $c = 0, 1728$, the results follow for the same reason from the multiplicity of the zero. □

Remark 40.3.10. The definition of $j(z)$ is now standard, but involves some choices. In some circumstances (including the generalization to abelian surfaces, see 43.5.7), it is more convenient to remember the values of the Eisenstein series themselves, as follows. To $z \in \mathcal{H}$, we associate the pair $(E_4(z), E_6(z)) \in \mathbb{C}^2$; if $\gamma \in \Gamma$ and $z' = \gamma z$, then

$$(E_4(z'), E_6(z')) = (\delta^4 E_4(z), \delta^6 E_6(z))$$

where $\delta = \jmath(\gamma; z) \in \mathbb{C}^\times$. We therefore define the **weighted projective** (4, 6)**-space** by

$$\mathbb{P}(4,6)(\mathbb{C}) := (\mathbb{C}^2 \smallsetminus \{(0,0)\})/\sim$$

where

$$(E_4, E_6) \sim (\delta^4 E_4, \delta^6 E_6)$$

for $\delta \in \mathbb{C}^\times$. We write equivalence classes $(E_4 : E_6) \in \mathbb{P}(4,6)(\mathbb{C})$. The map

$$\begin{aligned} j : \mathbb{P}(4,6)(\mathbb{C}) &\to \mathbb{P}^1(\mathbb{C}) \\ (E_4 : E_6) &\mapsto j(E_4 : E_6) = \frac{1728 E_4^3}{E_4^3 - E_6^2} \end{aligned}$$

is well-defined and bijective—see Silverman [Sil2009, Proposition III.1.4(b)].

To conclude, we give a complete description of the ring of (holomorphic) modular forms. By 40.2.3, the $\mathbb{C}$-vector space

$$M(\Gamma) = \bigoplus_{k \in 2\mathbb{Z}} M_k(\Gamma)$$

under multiplication has the structure of a (graded) $\mathbb{C}$-algebra; we call $M(\Gamma)$ the **ring of modular forms** for Γ.

Theorem 40.3.11. *We have $M(\Gamma) = \mathbb{C}[E_4, E_6]$, i.e., every modular form for $\Gamma = \mathrm{PSL}_2(\mathbb{Z})$ can be written as a polynomial in E_4, E_6.*

In particular, $M_k(\Gamma) = \{0\}$ for $k < 0$.

Proof. We have $M(\Gamma) \supseteq \mathbb{C}[E_4, E_6]$, so we prove the reverse inclusion. We refer to Proposition 40.3.4, and ask for solutions $a_1, a_2, a_3 \in \mathbb{Z}_{\geq 0}$ to $a_1 + a_2/2 + a_3/3 = k/12$. When $k < 0$, there are no such solutions; when $k = 0, 2, 4, 6, 8, 10$, there is a unique solution, and we find that $M_k(\Gamma)$ is spanned by $1, 0, E_4, E_6, E_4^2, E_4E_6$, respectively.

For all even $k \geq 4$, there exist $a, b \in \mathbb{Z}_{\geq 0}$ such that $4a + 6b = k$ (if $k \geq 10$ and $k \equiv 2 \pmod 4$, then $k - 6 \geq 4$ and $4 \mid k$), so $E_4^a E_6^b \in M_k(\Gamma)$ and $(E_4^a E_6^b)(\infty) = 1$ (by 40.2.22). Let $S_k(\Gamma) \subseteq M_k(\Gamma)$ be the subspace of forms that vanish at ∞. Then $M_k(\Gamma) = \mathbb{C}E_4^a E_6^b \oplus S_k(\Gamma)$ by linear algebra, and by the previous paragraph, $S_k(\Gamma) = \{0\}$ for $k \leq 10$.

We claim that multiplication by Δ furnishes an isomorphism $M_k(\Gamma) \xrightarrow{\sim} S_{k+12}(\Gamma)$ of $\mathbb{C}$-vector spaces for all k: division by Δ defines an inverse because Δ has a simple zero at ∞ by 40.3.6 and no zeros in $\mathbf{H}^2$. The result now follows by induction on $k \geq 0$. $\square$

More generally, one can study modular forms for congruence subgroups (section 35.4) of $\mathrm{PSL}_2(\mathbb{Z})$ in an explicit way, as the following example illustrates.

Example 40.3.12. At the end of section 35.4, we examined a fundamental domain for the group $\Gamma(2)$, defined by (35.4.8). As with $X(1)$, the homeomorphism (35.4.11) can be given by a holomorphic map

$$\lambda\colon X(2) \xrightarrow{\sim} \mathbb{P}^1(\mathbb{C})$$

obtained from Eisenstein series for $\Gamma(2)$, analogous to $j(z)$. The map λ satisfies $\lambda(\gamma z) = \lambda(z)$ for all $\gamma \in \Gamma(2)$ and in particular is invariant under $z \mapsto z + 2$. One can compute its Fourier expansion in terms of $q^{1/2} = e^{\pi i z}$ as:

$$\lambda(z) = 16q^{1/2} - 128q + 704q^{3/2} - 3072q^2 + 11488q^{5/2} - 38400q^3 + \dots. \quad (40.3.13)$$

Since $j(z)$ induces a degree $6 = [\Gamma(2) : \Gamma(1)]$ map $X(2) \to X(1)$, we find the relationship

$$j = 256\frac{(\lambda^2 - \lambda + 1)^3}{\lambda^2(\lambda - 1)^2}. \quad (40.3.14)$$

From (40.3.14) (and the first term), the complete series expansion (40.3.13) can be obtained recursively.

As a uniformizer for a congruence subgroup of $\mathrm{PSL}_2(\mathbb{Z})$, the function $\lambda(z)$ has a moduli interpretation (cf. 40.1.11): there is a family of elliptic curves over $X(2)$ equipped with extra structure. Specifically, given $\lambda \in \mathbb{P}^1(\mathbb{C}) \setminus \{0, 1, \infty\}$, the corresponding elliptic curve with extra structure is given by the **Legendre curve**

$$E_\lambda\colon y^2 = x(x - 1)(x - \lambda),$$

equipped with the isomorphism $(\mathbb{Z}/2\mathbb{Z})^2 \xrightarrow{\sim} E[2]$ determined by sending the standard generators to the 2-torsion points $(0, 0)$ and $(1, 0)$. The map j is the map that forgets this additional torsion structure on a Legendre curve and remembers only isomorphism class.

40.4 Theta series

An important class of classical modular forms arise via theta series, which count the number of representations of an integer by a positive definite quadratic form. We present only a small fraction of the general theory here.

Let $Q\colon \mathbb{Z}^m \to \mathbb{Z}$ be a positive definite integral quadratic form in m variables and suppose that $m = 2k$ is a positive even integer. We define the **theta series** of Q by

$$\Theta_Q : \mathbf{H}^2 \to \mathbb{C}$$

$$\Theta_Q(z) = \sum_{x \in \mathbb{Z}^m} e^{2\pi i Q(x) z} = \sum_{n=0}^{\infty} r_Q(n) q^n \quad (40.4.1)$$

where $q = e^{2\pi i z}$ and

$$r_Q(n) = \#\{x \in \mathbb{Z}^m : Q(x) = n\} < \infty$$

counts the number of lattice points on the sphere of radius $\sqrt{n}$.

Lemma 40.4.2. *$\Theta_Q(z)$ is a holomorphic function.*

Proof. Since Q is positive definite, there exists $c \in \mathbb{R}_{>0}$ such that

$$Q(x) \geq c(x_1^2 + \cdots + x_m^2).$$

Thus $r_Q(n) = O(n^k)$, and the series $\Theta_q(z)$ is majorized by (a constant multiple of) $\sum_{n=1}^{\infty} n^k q^n$, so converges to a holomorphic function. ◻

Let $[T]$ be the Gram matrix for the symmetric bilinear form associated to Q; then $[T] \in \mathrm{M}_m(\mathbb{Z})$ is an integral symmetric matrix with even diagonal entries. Let $d = \det Q = \det[T] \in \mathbb{Z}$. Then $dA^{-1} \in \mathrm{M}_m(\mathbb{Z})$ is the adjugate matrix: it is again symmetric.

Definition 40.4.3. The least positive integer $N \in \mathbb{Z}_{>0}$ such that NA^{-1} is integral with even diagonal entries is called the **level** of Q.

We recall the definition of the congruence subgroups 35.4.5.

Theorem 40.4.4. *The theta series $\Theta_Q(z)$ is a modular form of weight k for $\Gamma_1(N)$.*

Proof. Unfortunately, in this generality the proof would take us too far afield. Fundamentally, the transformation formula for Θ_Q follows from Poisson summation and careful computations: see Eichler [Eic73, §I.3, Proposition 2], Miyake [Miy2006, Corollary 4.9.5], or Ogg [Ogg69, Chapter VI]. ◻

40.4.5. We can be a bit more specific about the transformation group for $\Theta_Q(z)$ as follows. To Q, we associate the character χ defined by $\chi(n) = \left(\dfrac{(-1)^k \det Q}{n}\right)$. Then

$$\Theta_Q(\gamma z) = \chi(d)(cz + d)^k \Theta_Q(z) \qquad \text{for all } \gamma = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in \Gamma_0(N);$$

accordingly, we say Θ_Q is a modular form of **level** N **with character** χ.

40.5 Hecke operators

Significantly, the $\mathbb{C}$-vector space $M_k(\Gamma_0(N))$ of modular forms of weight k for $\Gamma_0(N)$ carries with it an action of commuting semisimple operators, called *Hecke operators*. These operators may be interpreted as averaging modular forms over sublattices of a fixed index; for efficiency, we work with a more explicit definition. For further reference, see e.g. Diamond–Shurman [DS2005, Chapter 5] or Miyake [Miy2006, §2.7, §4.5].

Throughout, let $N \in \mathbb{Z}_{\geq 1}$. Let

$$\mathcal{O} = \mathcal{O}_0(N) := \left\{ \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in \mathrm{M}_2(\mathbb{Z}) : N \mid c \right\} \subseteq \mathrm{M}_2(\mathbb{Z})$$

be the standard Eichler order of level N in $\mathrm{M}_2(\mathbb{Q})$, so that $\Gamma = \Gamma_0(N) = \mathcal{O}^1/\{\pm 1\}$.

Let $n \in \mathbb{Z}_{\geq 1}$ with $\gcd(n, N) = 1$. We consider the set of matrices

$$\mathcal{O}_n = \{\alpha \in \mathcal{O} : \det(\alpha) = n\}. \tag{40.5.1}$$

Visibly, there is a left (and right) action of $\mathcal{O}^1$ on $\mathcal{O}_n$ by multiplication.

Lemma 40.5.2. *A system of representatives of* $\mathcal{O}^1\backslash\mathcal{O}_n$ *is given by the set of matrices of the form* $\begin{pmatrix} a & b \\ 0 & d \end{pmatrix}$ *with* $ad = n$, $a > 0$, *and* $0 \leq b < d$.

Proof. The lemma follows as in Lemma 26.4.1(b) using the theory of elementary divisors, but applying row operations (acting on the left). ◻

Example 40.5.3. When $p \nmid N$ is prime, the set $\mathcal{O}^1\backslash\mathcal{O}_p$ is represented by the $p + 1$ matrices

$$\begin{pmatrix} p & 0 \\ 0 & 1 \end{pmatrix}, \begin{pmatrix} 1 & 0 \\ 0 & p \end{pmatrix}, \begin{pmatrix} 1 & 1 \\ 0 & p \end{pmatrix}, \dots, \begin{pmatrix} 1 & p-1 \\ 0 & p \end{pmatrix}.$$

Definition 40.5.4. For $n \in \mathbb{Z}_{\geq 1}$ with $\gcd(n, N) = 1$, we define the **Hecke operator**

$$\begin{aligned} T(n) : M_k(\Gamma) &\to M_k(\Gamma) \\ (T(n)f)(z) &= n^{k/2-1} \sum_{\mathcal{O}^1\alpha \in \mathcal{O}^1\backslash\mathcal{O}_n} \jmath(\alpha; z)^{-k} f(\alpha z). \end{aligned} \tag{40.5.5}$$

By the condition of automorphy $f(\gamma z) = \jmath(\gamma; z)^k f(z)$ and the cocycle relation (40.2.5), the Hecke operators are well-defined and preserve weight k invariance.

40.5.6. By Lemma 40.5.2, we have more explicitly

$$(T(n)f)(z) = n^{k-1} \sum_{\substack{ad=n \\ a>0}} \frac{1}{d^k} \sum_{b=0}^{d-1} f\left(\frac{az+b}{d}\right). \tag{40.5.7}$$

Accordingly, if $f(z) = \sum_{n=0}^{\infty} a_n q^n$, then $(T(n)f)(z) = \sum_{m=0}^{\infty} b_m q^m$ where

$$b_m = \sum_{\substack{d \mid \gcd(m,n) \\ d>0}} d^{k-1} a_{mn/d^2} \tag{40.5.8}$$

so in particular for $n = p$ prime we have

$$b_m = a_{pm} + \begin{cases} p^{k-1} a_{m/p}, & \text{if } p \mid m; \\ 0, & \text{if } p \nmid m. \end{cases} \tag{40.5.9}$$

Applying (40.5.9), we see that if $f \in M_k(\Gamma)$ has $f(\infty) = 0$ (equivalently, $a_0 = 0$), then the same is true for $T(n)f$. Repeating this for the functions $f[\gamma]_k$ with $\gamma \in \mathrm{PSL}_2(\mathbb{Z})$ (as in (40.2.14)), we conclude that the operators $T(n)$ act on the space of cusp forms $S_k(\Gamma) \subset M_k(\Gamma)$.

Proposition 40.5.10. *For $m, n \in \mathbb{Z}_{\geq 1}$, we have*

$$T(m)T(n) = \sum_{d \mid \gcd(m,n)} d^{k-1} T(mn/d^2). \tag{40.5.11}$$

In particular, if $\gcd(m, n) = 1$ *then*

$$T(m)T(n) = T(n)T(m) = T(mn)$$

and if p is prime and $r \geq 1$ then

$$T(p)T(p^r) = T(p^{r+1}) + p^{k-1} T(p^{r-1}).$$

Proof. These statements follow directly from the expansion (40.5.8). □

Theorem 40.5.12. *The Hecke operators $T(n)$ for* $\gcd(n, N) = 1$ *on $M_k(\Gamma_0(N))$ generate a commutative, semisimple $\mathbb{Z}$-algebra.*

Proof. See e.g. Diamond–Shurman [DS2005, Theorem 5.5.4]. Briefly, we treat Eisenstein series separately and work with cusp forms $S_k(\Gamma_0(N))$. To prove that the operators are semisimple, we would need to show that the **Petersson inner product**

$$\langle f, g \rangle = \int_{\Gamma \backslash \mathbf{H}^2} f(z)\overline{g(z)} y^k \, \mathrm{d}\mu(z)$$

is well-defined, positive, and nondegenerate, and then verify that the operators are normal with respect to this inner product. □

By Theorem 40.5.12 and linear algebra, there exists a $\mathbb{C}$-basis $f_i(z)$ of $M_k(\Gamma_0(N))$ consisting of simultaneous eigenfunctions for all $T(n)$.

Exercises

▷ 1. Let $f : U \to \mathbb{C}$ be a function that is meromorphic in an open neighborhood $U \supseteq \mathbb{C}$ with $z \in U$, and let C be a contour along an arc of a circle of radius $\epsilon > 0$ centered at z contained in U with total angle θ. Show that

$$\lim_{\epsilon \to 0} \int_C \frac{\mathrm{d}f}{f} = \theta i \operatorname{ord}_z(f).$$

▷ 2. Prove the formula

$$\pi \cot(\pi z) = \sum_{m=-\infty}^{\infty} \frac{1}{z+m}$$

for $z \in \mathbb{C}$. *[Hint: the difference $h(z)$ of the left- and right-hand sides is bounded away from $\mathbb{Z}$, invariant under $z \mapsto T(z) = z + 1$, and in a neighborhood of* 0 *is holomorphic (both sides have principal part $1/z$) so bounded; thus $h(z)$ is bounded in $\mathbb{C}$ and hence constant.]*

▹ 3. In this exercise, we give Euler's evaluation of $\zeta(k)$ in terms of Bernoulli numbers. As in Exercise 36.12, define the series

$$\frac{x}{e^x - 1} = \sum_{k=0}^{\infty} B_k \frac{x^k}{k!} = 1 - \frac{x}{2} + \frac{1}{6}\frac{x^2}{2!} - \frac{1}{30}\frac{x^4}{4!} + \dots \in \mathbb{Q}[[x]]. \tag{40.5.13}$$

(a) Plug in $x = 2iz$ into (40.5.13) to obtain

$$z \cot z = 1 + \sum_{k=2}^{\infty} B_k \frac{(2iz)^k}{k!}.$$

(b) Take the logarithmic derivative of

$$\sin z = z \prod_{n=1}^{\infty} \left(1 - \frac{z^2}{n^2\pi^2}\right)$$

to show

$$z \cot z = 1 - 2 \sum_{\substack{k=2 \\ k \text{ even}}}^{\infty} \sum_{n=1}^{\infty} \left(\frac{z}{n\pi}\right)^k.$$

(c) Conclude that

$$\zeta(k) = \sum_{n=1}^{\infty} \frac{1}{n^k} = -\frac{1}{2}\frac{(2\pi i)^k}{k!} B_k$$

for $k \in 2\mathbb{Z}_{\geq 1}$.

4. We defined Eisenstein series $G_k(z)$ for $k \geq 4$, and found $G_k(z) \in \mathrm{M}_k(\mathrm{SL}_2(\mathbb{Z}))$ are modular forms of weight k for $\mathrm{SL}_2(\mathbb{Z})$. The case $k = 2$ is also important, though we must be a bit more careful in its analysis. Let

$$G_2(z) := \sum_{c \in \mathbb{Z}} \sum_{\substack{d \in \mathbb{Z} \\ (c,d) \neq (0,0)}} \frac{1}{(cz + d)^2}.$$

(a) Show that $G_2(z)$ converges (conditionally) and satisfies

$$G_2(z) = 2\zeta(2) - 8\pi^2 \sum_{n=1}^{\infty} \sigma(n) q^n$$

where $q = e^{2\pi i z}$.

(b) Show that $G_2(z + 1) = G_2(z)$ and

$$G_2\left(\frac{-1}{z}\right) = G_2(z) - \frac{2\pi i}{z}.$$

[Hint: use a telescoping series and rearrange terms.]

(c) Conclude that

$$G_2(\gamma z) = G_2(z) - \frac{2\pi i c}{cz+d} \qquad \text{for all } \gamma = \begin{pmatrix} a & b \\ c & d \end{pmatrix}.$$

[Hint: use the cocycle relation.]

(d) Define

$$G_2^*(z) = G_2(z) - \frac{\pi}{\operatorname{Im} z}.$$

Show that $G_2^*(z)$ is weight 2 invariant under $\mathrm{SL}_2(\mathbb{Z})$.

5. In this exercise, we give a proof using modular forms of the formula for the number of ways of representing an integer as the sum of four squares, due to Jacobi. Consider the function

$$\vartheta(z) := \sum_{n=-\infty}^{\infty} q^{n^2}$$

where $q := e^{2\pi i z}$ and $z \in \mathbf{H}^2$. [The function $\vartheta(z)$ is a theta series for the univariate quadratic form $x \mapsto x^2$.] Let $r_4(n)$ be the number of ways of representing $n \geq 0$ as the sum of 4 squares.

(a) Show that

$$\Theta_Q(z) := \vartheta(z)^4 = 1 + \sum_{n=1}^{\infty} r_4(n) q^n = 1 + 8q + 12q^2 + \dots.$$

(b) Show that $\Theta_Q(q) \in \mathrm{M}_2(\Gamma_0(4))$ is a modular form of weight 2 on $\Gamma_0(4)$.

(c) Show that $\dim_{\mathbb{C}} M_2(\Gamma_0(4)) = 2$ and $\dim_{\mathbb{C}} S_2(\Gamma_0(4)) = 0$.

(d) Let

$$G_{2,2}(z) = G_2(z) - 2G_2(2z)$$
$$G_{2,4}(z) = G_2(z) - 4G_2(4z).$$

Show that $G_{2,2}, G_{2,4}$ are a basis for $\mathrm{M}_2(\Gamma_0(4))$. *[Hint: use Exercise 40.4(c).]*

(e) Show that

$$E_{2,2}(z) := -\frac{3}{\pi^2} G_{2,2}(z) = 1 + 24 \sum_{n=1}^{\infty} \sigma^{(2)}(n) q^n$$
$$E_{2,4}(z) := -\frac{1}{\pi^2} G_{2,4}(z) = 1 + 8 \sum_{n=1}^{\infty} \sigma^{(4)}(n) q^n$$

where

$$\sigma^{(m)}(n) = \sum_{m \nmid d \mid n} d.$$

(f) Matching the first few coefficients, show that

$$\Theta_Q(z) = E_{2,4}(z).$$

Conclude that $r_4(n) = 8\sigma^{(4)}(n)$ for all $n > 0$.

This chapter is licensed under the terms of the Creative Commons Attribution-NonCommercial 4.0 International License (http://creativecommons.org/licenses/by-nc/4.0/), which permits any noncommercial use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license and indicate if changes were made.

The images or other third party material in this chapter are included in the chapter's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the chapter's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.

Chapter 41

Brandt matrices

In this chapter, we revisit classes of quaternion ideals: organizing ideals of given norm in terms of their classes, we find modular forms.

41.1 ▹ Brandt matrices, neighbors, and modular forms

Let B be a quaternion algebra over $\mathbb{Q}$. A major theme of this text has been the study of classes of quaternion ideals, beginning with chapter 17. When B is indefinite, we saw (Theorem 17.8.3, treated broadly in chapter 28) that strong approximation applies: via the reduced norm, very often the conclusion is that the class set is trivial.

We are left with the case that B is definite. By the geometry of numbers (see section 17.5) we found that the number of ideal classes of an order is finite, generated by ideals of small reduced norm. (Studying the zeta function we found a mass formula in chapter 25, and then studying quadratic embeddings we found a class number formula in section 30.8.) We now pursue this further: there is an exquisite arithmetic and combinatorial structure to be found by counting right ideals of given norm by their classes as follows. We begin in this section by an introduction and survey (working over $\mathbb{Q}$).

Let $O \subset B$ be an order. Let $\operatorname{Cls} O$ be the right class set of O, keeping track of the isomorphism classes of invertible right O-ideals in B. Let $h := \#\operatorname{Cls} O$ be the (right) class number of O, and let $I_1, \dots, I_h \subseteq B$ represent the distinct classes in $\operatorname{Cls} O$.

Let $n \in \mathbb{Z}_{\geq 1}$. We define an $h \times h$-matrix $T(n) \in \mathrm{M}_h(\mathbb{Z})$ with nonnegative integer entries, called the n-**Brandt matrix**, by

$$\begin{aligned} T(n)_{ij} &:= \#\{J \subset I_j : \operatorname{nrd}(J) = n \operatorname{nrd}(I_j) \text{ and } [J] = [I_i]\} \\ &= \#\{J \subset I_j : [I_j : J] = n^2 \text{ and } [J] = [I_i]\}. \end{aligned} \tag{41.1.1}$$

(The notation $T(n)$ deliberately overloads that of the Hecke operators defined in section 40.5: keep reading to see why!) The Brandt matrix $T(n)$ depends on O, but for brevity we do not include this in the notation. In the jth column of the Brandt matrix $T(n)$, we look at the subideals of I_j with index n^2 and count them in the ijth entry according to the class $[I_i]$ they belong to. If $n = p$ is prime and $p \nmid N = \operatorname{disc} O$, then there are

© The Author(s) 2021

J. Voight, *Quaternion Algebras*, Graduate Texts in Mathematics 288,
https://doi.org/10.1007/978-3-030-56694-4_41

exactly $p+1$ such ideals, so the sum of the entries in every column in $T(p)$ is equal to $p+1$.

Example 41.1.2. We continue with Example 17.6.3. We have $B = \left(\dfrac{-1,-23}{\mathbb{Q}}\right)$ of discriminant 23 and a maximal order O with three ideal classes $[I_1], [I_2], [I_3]$. In (17.6.5), we found three ideals in $I_1 = O$: two belong to the class $[I_2]$ and the third is principal, belonging to $[I_1]$. This gives the first column of the matrix as $(1,2,0)^{\mathsf{t}}$. Computing further, we find

$$T(2) = \begin{pmatrix} 1 & 1 & 0 \\ 2 & 1 & 3 \\ 0 & 1 & 0 \end{pmatrix}.$$

In a similar manner, we compute

$$T(3) = \begin{pmatrix} 0 & 1 & 3 \\ 2 & 3 & 0 \\ 2 & 0 & 1 \end{pmatrix}, \quad T(101) = \begin{pmatrix} 30 & 28 & 24 \\ 56 & 54 & 60 \\ 16 & 20 & 18 \end{pmatrix}.$$

41.1.3. There is a second and computationally more efficient way to define the Brandt matrix using representation numbers of quadratic forms. Let $q_i = \operatorname{nrd}(I_i)$, let $O_i = O_{\mathsf{L}}(I_i)$, and let $w_i = \#O_i^\times/\{\pm 1\} < \infty$. Then

$$T(n)_{ij} = \frac{1}{2w_i}\#\{\alpha \in I_j I_i^{-1} : \operatorname{nrd}(\alpha) q_i/q_j = n\} :$$

indeed, $\alpha I_i = J \subseteq I_j$ with $\operatorname{nrd}(J) = n \operatorname{nrd}(I_j)$ if and only if $\alpha \in I_j I_i^{-1} = (I_j : I_i)_{\mathsf{L}}$ and $\operatorname{nrd}(\alpha) q_i = p q_j$, and α is well defined up to right multiplication by $\mu \in O_i^\times$. Now

$$\begin{aligned} Q_{ij} \colon I_j I_i^{-1} &\to \mathbb{Z} \\ Q_{ij}(\alpha) &= \operatorname{nrd}(\alpha)\frac{q_i}{q_j} \end{aligned} \tag{41.1.4}$$

is a positive definite quadratic form, so it suffices to enumerate lattice points!

Example 41.1.5. Returning to our example, we have

$$T(n)_{ii} = \frac{1}{2w_i}\#\{\gamma \in O_i : \operatorname{nrd}(\gamma) = n\}.$$

For $i = 1$, we have $w_1 = 2$ and with $\gamma = t + x\alpha + y\beta + z\alpha\beta$ and $t, x, y, z \in \mathbb{Z}$, by (17.6.6)

$$\operatorname{nrd}(\gamma) = t^2 + ty + x^2 + xz + 6y^2 + 6z^2$$

so $T(p)_{11}$ counts half the number of representations of n by this positive definite quaternary quadratic form.

There is a third way to understand Brandt matrices which is visual and combinatorial.

Definition 41.1.6. Let $I, J \subseteq O$ be invertible right O-ideals. We say J is a **n-neighbor** of I if $J \subseteq I$ and $n \operatorname{nrd}(I) = \operatorname{nrd}(J)$.

The **n-Brandt graph** is the directed graph with vertices $\operatorname{Cls} O$ and a directed edge from $[I_i]$ to $[J]$ for each n-neighbor $J \subseteq I_i$.

There is no extra content here, just a reinterpretation: the n-Brandt matrix is simply the *adjacency* matrix of the n-Brandt graph.

Example 41.1.7. Returning a third time to our example, we have the 2-Brandt graph, as in Figure 41.1.8.

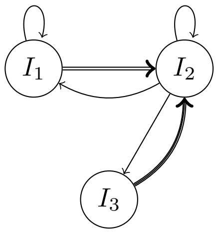


Figure 41.1.8: The 2-Brandt graph for discriminant 23

41.1.9. For $n = p$ prime, there is another way to think of the p-Brandt graph. Consider the directed graph whose vertices are invertible right O-ideals whose reduced norm is a power of p, and draw a directed edge from I to J if J is a p-neighbor of I. If $p \nmid N$, then this graph is a $(p+1)$-regular directed tree, that is to say, from each vertex there are $p+1$ directed edges. The notion of belonging to the same ideal class induces an equivalence relation on this graph, and the quotient is the p-Brandt graph.

It is helpful to think of the matrices $T(n)$ as operators on a space, so we define the **Brandt module** $M_2(O)$ to be the $\mathbb{C}$-vector space with basis $\operatorname{Cls} O$ and equipped with the action of Brandt matrices $T(n)$ for $n \in \mathbb{Z}_{\geq 0}$ on the right.

The Brandt matrices have two important properties. First, they commute: by a quaternionic version of the Sun Zi theorem (CRT), if $\gcd(m, n) = 1$ then

$$T(m)T(n) = T(n)T(m). \tag{41.1.10}$$

Informally, we might say that the process of taking m-neighbors commutes with the process of taking n-neighbors, when m, n are coprime. Second, they are self-adjoint for the pairing on $M_2(O)$ given by

$$\langle [I_i], [I_j] \rangle = \begin{cases} 1/w_i, & \text{if } i = j; \\ 0, & \text{else.} \end{cases}$$

The proof of self-adjointness is contained in the equality $w_i T(n)_{ij} = w_j T(n)_{ji}$, and this follows from a bijection induced by the standard involution.

Therefore the matrices $T(n)$ are semisimple (diagonalizable) and $M_2(O)$ has a simultaneous basis of eigenvectors, which we call a **eigenbasis**. The row $e = (1, 1, \dots, 1)$ is always an eigenvector (by the sum of columns) with eigenvalue $a_p(e) = p + 1$ for $p \nmid N$.

Example 41.1.11. We check that $T(2)T(3) = T(3)T(2)$ from Example 41.1.2; and $w_1, w_2, e_3 = 2, 1, 3$, so we verify that

$$\begin{pmatrix} 2 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 3 \end{pmatrix} T(2) = \begin{pmatrix} 2 & 2 & 0 \\ 2 & 1 & 3 \\ 0 & 3 & 0 \end{pmatrix}$$

is symmetric. The characteristic polynomial of $T(2)$ is $(x-3)(x^2+x-1)$, and for $T(3)$ it is $(x-4)(x^2-5)$. We find the eigenbasis $e = (1,1,1)$ and $f_\pm = (4, \pm\sqrt{5}-3, \mp 3\sqrt{5}+3)$, and observe the required orthogonality

$$\langle e, f_\pm \rangle = \frac{1}{2}\cdot 4 + (\pm\sqrt{5}-3) + \frac{1}{3}(\mp 3\sqrt{5}+3) = 0.$$

By now, hopefully the reader is convinced that the Brandt matrices capture interesting arithmetic information about the order O and that they are not difficult to compute.

Now comes the modular forms: the second way of viewing Brandt matrices shows that we should be thinking of a generating series for the representation numbers of the quadratic forms Q_{ij} defined in (41.1.4). As in section 40.4, we define the **theta series** for the quadratic form Q_{ij}

$$\Theta_{ij}(q) := \sum_{n=0}^{\infty} T(n)_{ij} q^n = \frac{1}{2w_i} \sum_{\gamma \in I_j I_i^{-1}} q^{Q_{ij}(\gamma)} \in \mathbb{Z}[[q]].$$

Letting $q := e^{2\pi i z}$ for $z \in \mathbf{H}^2$, by Theorem 40.4.4 (a consequence of Poisson summation), the function $\Theta_{ij}(z) : \mathbf{H}^2 \to \mathbb{C}$ is a modular form of weight 2 for an explicit congruence subgroup of $\mathrm{SL}_2(\mathbb{Z})$: for example, if O is an Eichler order with reduced discriminant N, then $\theta_{ij}(q) \in M_2(\Gamma_0(N))$ (trivial character).

There is something enduringly magical about the fact that the entries of Brandt matrices (arithmetic) give Fourier coefficients of holomorphic modular forms (geometric, analytic).

Example 41.1.12. Returning one last time to our example, the space $M_2(\Gamma_0(23))$ of modular forms of weight 2 and level $\Gamma_0(23)$ has eigenbasis e_{23}, f_+, f_- where

$$e_{23}(z) = \frac{11}{12} + \sum_{n=1}^{\infty} \sigma^*(n) q^n, \qquad \sigma^*(n) = \sum_{\substack{d \mid n \\ 23 \nmid d}} d$$

and

$$f_\pm(z) := q - \frac{\pm\sqrt{5}+1}{2} q^2 + \sqrt{5} q^3 + \dots$$

are cusp forms matching the eigenbasis in Example 41.1.11.

One of the main applications of Brandt matrices is to express the trace of the Hecke operator in terms of arithmetic data, as follows.

Theorem 41.1.13. *Let B be a definite quaternion algebra over $\mathbb{Q}$ of discriminant D and $O \subseteq B$ a maximal order. For $d < 0$ a fundamental discriminant, define*

$$h_D(d) := \frac{h(d)}{w_d} \prod_{p \mid D} \left(1 - \left(\frac{d}{p}\right)\right)$$

where $h(d)$ is the class number of $\mathbb{Q}(\sqrt{d})$, w_d its number of roots of unity, and $\left(\frac{d}{p}\right)$ is the Kronecker symbol. For $d' < 0$ a discriminant with $d' = df^2$ and d fundamental, define $h_D(d') = h_D(d)$. Then the trace of the nth Brandt matrix associated to O is

$$\operatorname{tr} T(n) = \sum_{\substack{t \in \mathbb{Z} \\ t^2 < 4n}} h_D(t^2 - 4n) + \begin{cases} \varphi(D)/12, & \text{if } n \text{ is a square;} \\ 0, & \text{otherwise} \end{cases}$$

where φ is the Euler totient function.

Theorem 41.1.13 is a special case of Main Theorem 41.5.2, see Example 41.5.8. In a surprising way, it exhibits a relationship between traces of Hecke operators and (modified) class numbers of imaginary quadratic fields!

41.2 Brandt matrices

To begin, we define the all-important Brandt matrices in the generality considered in this text.

Let R be a global ring with eligible set $S \subseteq \operatorname{Pl} F$, let B be an S-definite quaternion algebra over F of discriminant $\operatorname{disc} B = \mathfrak{D}$ and let $O \subset B$ be an R-order in B with reduced discriminant $\operatorname{discrd} O = \mathfrak{N}$. The reader will probably have in mind the case where F is a totally real (number) field, S the set of real (archimedean) places of F, and B a definite quaternion algebra over F; but the arguments hold just as well with a larger set S of ramified places or in the function field case.

41.2.1. Let $\operatorname{Cls} O$ be the right class set of O. By Corollary 27.6.20, the class number $h = \#\operatorname{Cls} O < \infty$ is finite. Let $I_1, \dots, I_h$ be a set of representative invertible right O-ideals for $\operatorname{Cls} O$. For $i = 1, \dots, h$, let $O_i = O_{\mathsf{L}}(I_i)$ be the left order of I_i; then O_i depends on the choice of I_i but its isomorphism class (i.e., type) is independent of the choice of I_i. (Due to the possible presence of two-sided ideals, there may be repetition of types among the orders O_i.)

41.2.2. Let $\mathfrak{n} \subset R$ be a nonzero ideal. For each j, we consider the set of right invertible O-ideals $J \subseteq I_j$ with $\operatorname{nrd}(J) = \mathfrak{n} \operatorname{nrd}(I_j)$, and we count them according to their class in $\operatorname{Cls} O$:

$$T(\mathfrak{n})_{ij} := \#\{J \subseteq I_j : \operatorname{nrd}(J) = \mathfrak{n} \operatorname{nrd}(I_j) \text{ and } [J] = [I_i]\} \in \mathbb{Z}_{\geq 0}. \tag{41.2.3}$$

A containment $J \subseteq I_j$ of right O-ideals yields a compatible product $J' = JI_j^{-1}$ and thus an invertible right O_j-ideal with reduced norm $\operatorname{nrd}(J') = \operatorname{nrd}(JI_j^{-1}) = \mathfrak{n}$, and

conversely. So equivalently

$$T(\mathfrak{n})_{ij} = \#\{J' \subseteq O_j : \operatorname{nrd}(J') = \mathfrak{n} \text{ and } [J'I_j] = [I_i]\}.$$

(We could also rewrite $[J'I_j] = [I_i]$ in terms of classes of right ideals of O_j.)

Definition 41.2.4. The $\mathfrak{n}$-**Brandt matrix** for O is the matrix $T(\mathfrak{n}) \in \mathrm{M}_h(\mathbb{Z})$ whose (i,j)th entry is equal to $T(\mathfrak{n})_{i,j}$.

41.2.5. To make the definition more canonical, we define

$$M_2(O) := \operatorname{Map}(\operatorname{Cls} O, \mathbb{Z})$$

to be the set of maps from $\operatorname{Cls} O$ to $\mathbb{Z}$ (as sets). Then $M_2(O)$ has the structure of an abelian group under addition of maps, and it is a free $\mathbb{Z}$-module on the characteristic functions for $\operatorname{Cls} O$. The $\mathfrak{n}$-**Hecke operator** is defined to be

$$\begin{aligned} T(\mathfrak{n})\colon M_2(O) &\to M_2(O) \\ (T(\mathfrak{n})f)([I]) &= \sum_{\substack{J \subseteq I \\ \operatorname{nrd}(J) = \mathfrak{n}\operatorname{nrd}(I)}} f([J]) \end{aligned} \tag{41.2.6}$$

again the sum over all invertible right O-ideals $J \subseteq I$ with condition on the reduced norm. Visibly, this definition does not depend on the choice of representative I in its right ideal class. And in the basis of characteristic functions for I_i, the matrix of $T(\mathfrak{n})$ is precisely the $\mathfrak{n}$-Brandt matrix.

Brandt matrices may be given in terms of elements instead of ideals. Let $w_i = [O_i^\times : R^\times]$. By Proposition 32.3.7, since B is S-definite, the unit index $w_i < \infty$ is finite.

Lemma 41.2.7. *Let $\mathfrak{n}_{ij} = \mathfrak{n}\operatorname{nrd}(I_j)/\operatorname{nrd}(I_i)$ for $i,j = 1,\dots,h$. Then following statements hold.*

(a) *We have*

$$\begin{aligned} T(\mathfrak{n})_{ij} &= \#\left\{\alpha \in I_j I_i^{-1} : \operatorname{nrd}(\alpha)R = \mathfrak{n}_{ij}\right\}/O_i^\times \\ &= \frac{1}{w_i}\#\left\{\alpha \in I_j I_i^{-1} : \operatorname{nrd}(\alpha)R = \mathfrak{n}_{ij}\right\}/R^\times \end{aligned} \tag{41.2.8}$$

where we count orbits under right multiplication by $O_i^\times$ and $R^\times$, respectively.

(b) *If the class of $\mathfrak{n}_{ij}$ in $\mathrm{Cl}^+ R$ is nontrivial, then $T(\mathfrak{n})_{ij} = 0$.*

(c) *Suppose that $\mathfrak{n}_{ij} = n_{ij}R$ with $n_{ij} \in F_{>0}^\times$ totally positive. Then*

$$T(\mathfrak{n})_{ij} = \frac{1}{2w_i}\sum_{uR^{\times 2} \in R_{>0}^\times/R^{\times 2}} \#\left\{\alpha \in I_j I_i^{-1} : \operatorname{nrd}(\alpha) = u n_{ij}\right\} \tag{41.2.9}$$

where the sum is over a choice of representatives of totally positive units of R modulo squares.

Proof. We claim that there is a bijection

$$\begin{aligned}&\{J \subseteq I_j : \operatorname{nrd}(J) = \mathfrak{n}\operatorname{nrd}(I_j) \text{and} [J] = [I_i]\} \\ &\quad\leftrightarrow \left\{\alpha \in I_j I_i^{-1} : \operatorname{nrd}(\alpha) = \mathfrak{n}_{ij}\right\} / O_i^\times\end{aligned} \tag{41.2.10}$$

with orbits under right multiplication by $O_i^\times = O_\mathsf{R}(I_i^{-1})^\times$. Indeed, a containment $J \subseteq I_j$ of invertible right O-ideals with $[I_i] = [J]$ corresponds to $\alpha \in B^\times$ such that $\alpha I_i = J$, so in fact $\alpha \in (J : I_i)_\mathsf{L} = JI_i^{-1}$, and $\operatorname{nrd}(J) = \mathfrak{n}\operatorname{nrd}(I_j)$ translates into $\operatorname{nrd}(\alpha)\operatorname{nrd}(I_i) = \operatorname{nrd}(J) = \mathfrak{n}\operatorname{nrd}(I_j)$ or $\operatorname{nrd}(\alpha)R = \mathfrak{n}_{ij}$. Writing $JI_i^{-1} = \alpha O_i$, we see that α is unique up to multiplication on the right by $O_i^\times$. To finish (a), we note that the right action by $O_i^\times$ is free; and (b) follows from (a), since $\operatorname{nrd}(B^\times) \leq F_{>0}^\times$ as B is S-definite.

For (c), we just need to organize our generators; the sum in (41.2.9) is finite by the Dirichlet S-unit theorem. If $\operatorname{nrd}(\alpha)R = \mathfrak{n}_{ij}$ then $\operatorname{nrd}(\alpha) = un_{ij}$ for some $u \in R_{>0}^\times$. Multiplying by an element of $R_{>0}^\times$, we may suppose that $\operatorname{nrd}(\alpha)/n_{ij} = u$ belongs in a set of representatives for $R_{>0}^\times/R^{\times 2}$, and for $v \in R^\times$, we have $\operatorname{nrd}(v\alpha) = v^2\operatorname{nrd}(\alpha) = \operatorname{nrd}(\alpha)$ if and only if $v = \pm 1$, which gives us an extra factor 2. □

41.2.11. The advantage of (41.2.9) is that it can be expressed simply in terms of a quadratic form. Suppose that F is a number field and $R = \mathbb{Z}_F$, when this observation is especially clean. Since B is totally definite, as in 17.7.10, the space $B \hookrightarrow B \otimes_\mathbb{Q} \mathbb{R} \cong \mathbb{H}^n \cong \mathbb{R}^{4n}$ comes equipped with the positive definite quadratic form $Q = \operatorname{Tr}_{F/\mathbb{Q}}\operatorname{nrd}\colon B \to \mathbb{R}$, and if J is a R-lattice, then $J \cong \mathbb{Z}^{4n}$ embeds as a Euclidean lattice $J \hookrightarrow \mathbb{R}^{4n}$ with respect to this quadratic form. Therefore,

$$\{\alpha \in I_j I_i^{-1} : \operatorname{nrd}(\alpha) = un_{ij}\} \subseteq \{\alpha \in I_j I_i^{-1} : Q(\alpha) = \operatorname{Tr}_{F/\mathbb{Q}} un_{ij}\}$$

where the latter set is finite and effectively computable.

Finally, Brandt matrices are adjacency matrices.

Definition 41.2.12. Let $I, J \subseteq O$ be invertible right O-ideals. We say J is a $\mathfrak{n}$-**neighbor** of I if $J \subseteq I$ and $\operatorname{nrd}(J) = \mathfrak{n}\operatorname{nrd}(I)$.

The $\mathfrak{n}$-**Brandt graph** is the directed graph with vertices $\operatorname{Cls} O$ and a directed edge from $[I_i]$ to $[J]$ for each $\mathfrak{n}$-neighbor $J \subseteq I$.

By definition, the adjacency matrix of the $\mathfrak{n}$-Brandt graph is the $\mathfrak{n}$-Brant matrix $T(\mathfrak{n})$.

41.2.13. Let $\mathfrak{p} \nmid \mathfrak{N}$ be prime and suppose that the class of $\mathfrak{p}$ generates $\operatorname{Cl}_{G(O)} R$. Then by Proposition 28.5.18, we may take the ideals I_i to have reduced norm a power of $\mathfrak{p}$. Consider the directed graph whose vertices are the right O-ideals whose reduced norm is a power of $\mathfrak{p}$ with directed edges for each $\mathfrak{p}$-neighbor relation. (This graph is a regular directed tree by Proposition 41.3.1 below, every vertex has out degree equal to $\mathsf{N}\mathfrak{p} + 1$.) The equivalence relation of belonging to the same right ideal class (left equivalent by an element of $B^\times$) respects edges, and the quotient by this equivalence relation is the $\mathfrak{p}$-Brandt graph.

Remark 41.2.14. The Brandt graphs have interesting graph theoretic properties: they are *Ramanujan graphs* (also called *expander graph*), having high connectivity and are potentially useful in communication networks. In the simplest case where $F = \mathbb{Q}$ and B is the quaternion algebra of discriminant p, they were first studied by Ihara, then studied in specific detail by Lubotzky–Phillips–Sarnak [LPS88] and Margulis [Marg88]; for further reading, see the books by Lubotzky [Lub2010] and Sarnak [Sar90]. Over totally real fields, see work of Livné [Liv2001] as well as Charles–Goren–Lauter [CGL2009]. The proof that Brandt graphs are Ramanujan relies on the Ramanujan–Petersson conjecture, a deep statement proven by Deligne [Del74], giving bounds on coefficients of modular forms.

Remark 41.2.15. The space of functions on $\operatorname{Cls} O$ can itself be understood as a space of modular forms, a special case of the theory of *algebraic modular forms* due to Gross [Gro99]. This general formulation harmonizes with the double coset description given in section 38.7, via the canonical bijection $\operatorname{Cls} O \leftrightarrow B^\times\backslash\widehat{B}^\times/\widehat{O}^\times$, but without the geometry!

41.3 Commutativity of Brandt matrices

In this section, we examine basic properties of Brandt matrices—including that they commute.

Proposition 41.3.1. *The following statements hold.*

(a) *The sum of the entries* $\sum_i T(\mathfrak{n})_{ij}$ *in every column of* $T(\mathfrak{n})$ *is constant; if* $\mathfrak{n}$ *is coprime to* $\mathfrak{N}$*, then this constant is equal to* $\sum_{\mathfrak{d}\mid\mathfrak{n}} \mathsf{N}(\mathfrak{d})$*, where* $\mathsf{N}(\mathfrak{a}) = \#(R/\mathfrak{a})$ *is the absolute norm.*

(b) *If* $\mathfrak{m}, \mathfrak{n}$ *are relatively prime, then*

$$T(\mathfrak{m}\mathfrak{n}) = T(\mathfrak{m})T(\mathfrak{n}) = T(\mathfrak{n})T(\mathfrak{m}). \tag{41.3.2}$$

Proof. First we prove (a). The orders O_j are locally isomorphic, so by the local-global dictionary for lattices (Theorem 9.4.9), the number of invertible right O_j-ideals with given reduced norm is independent of j, giving the first statement. For the second statement, under the hypothesis that $\mathfrak{n}$ is coprime to $\mathfrak{N}$, for all $\mathfrak{p} \mid \mathfrak{n}$ we have $O_\mathfrak{p} \simeq \mathrm{M}_2(R_\mathfrak{p})$, and we counted right ideals in our pursuit of the zeta function: by Proposition 26.3.9, these counts are multiplicative, and by Lemma 26.4.1(b), the number of reduced norm $\mathfrak{p}^e$ is $1 + q + \dots + q^e$ where $q = \mathsf{N}(\mathfrak{p})$.

Statement (b) follows similar logic but with "unique factorization" of right ideals. As above, an invertible right O_j-ideal of reduced norm $\mathfrak{m}\mathfrak{n}$ by Lemma 26.3.6 factors uniquely into a compatible product of invertible lattices of reduced norm $\mathfrak{m}$ and $\mathfrak{n}$: organizing by classes, this says precisely that

$$T(\mathfrak{m}\mathfrak{n})_{ij} = \sum_{k=1}^{h} T(\mathfrak{m})_{ik}T(\mathfrak{n})_{kj} \tag{41.3.3}$$

which gives the matrix product $T(\mathfrak{m}\mathfrak{n}) = T(\mathfrak{m})T(\mathfrak{n})$. Repeating the argument interchanging the roles of $\mathfrak{m}$ and $\mathfrak{n}$, the result is proven. ∎

For prime powers coprime to $\mathfrak{N}$, we have a recursion for the $\mathfrak{p}^r$-Brandt matrices that is a bit complicated: the uniqueness of factorization fails when the product is a two-sided ideal, so we must account for this extra term. To this end, we need to keep track of the effect of multiplication by right ideals of R on the class set.

41.3.4. For an ideal $\mathfrak{a} \subseteq R$, let $P(\mathfrak{a}) \in \mathrm{M}_h(\mathbb{Z})$ be the permutation matrix given by $I_i \mapsto \mathfrak{a}I_i$. In other words, we place a 1 in the (i, j)th entry according as $[\mathfrak{a}I_j] = [I_i]$ (with 0 elsewhere). The matrix $P(\mathfrak{a})$ only depends on the class $[\mathfrak{a}] \in \mathrm{Cl}\,R$: in particular, if $\mathfrak{a}$ is principal then $P(\mathfrak{a})$ is the identity matrix. Therefore we have a homomorphism

$$\begin{aligned} P : \mathrm{Cl}\,R &\to \mathrm{GL}_h(\mathbb{Z}) \\ [\mathfrak{a}] &\mapsto P(\mathfrak{a}). \end{aligned}$$

We have

$$P(\mathfrak{a}\mathfrak{b}) = P(\mathfrak{a})P(\mathfrak{b}) = P(\mathfrak{b})P(\mathfrak{a})$$

and in particular $P(\mathfrak{a})P(\mathfrak{a}^{-1}) = 1$ and the image $P(\mathrm{Cl}\,R) \subseteq \mathrm{GL}_h(\mathbb{Z})$ is an abelian subgroup; however, this map need not be injective. Moreover, for all $\mathfrak{a}, \mathfrak{n}$ we have

$$P(\mathfrak{a})T(\mathfrak{n}) = T(\mathfrak{n})P(\mathfrak{a}) \tag{41.3.5}$$

by commutativity of multiplication by $\mathfrak{a}$.

As in 26.4.3, we say an integral right O-ideal I is **primitive** if we cannot write $I = \mathfrak{a}I'$ with I' integral and $\mathfrak{a} \subsetneq R$.

Proposition 41.3.6. *Let $\mathfrak{p} \nmid \mathfrak{N}$ be prime. Then for $r, s \in \mathbb{Z}_{\geq 0}$,*

$$T(\mathfrak{p}^r)T(\mathfrak{p}^s) = \sum_{i=0}^{\min(r,s)} \mathsf{N}(\mathfrak{p})^i T(\mathfrak{p}^{r+s-2i})P(\mathfrak{p})^i. \tag{41.3.7}$$

In particular, for all $r \geq 0$,

$$T(\mathfrak{p}^{r+2}) = T(\mathfrak{p}^{r+1})T(\mathfrak{p}) - \mathsf{N}(\mathfrak{p})T(\mathfrak{p}^r)P(\mathfrak{p}). \tag{41.3.8}$$

Proof. When $s = 0$, the matrix $T(1)$ is the identity and the result holds. We next consider the case $s = 1$, and will then proceed by induction, and consider the product $T(\mathfrak{p}^r)T(\mathfrak{p})$: its ijth entry

$$(T(\mathfrak{p}^r)T(\mathfrak{p}))_{ij} = \sum_{k=1}^{h} T(\mathfrak{p}^r)_{ik}T(\mathfrak{p})_{kj}$$

counts the number of compatible products of right ideals $J_r'J'$ where J_r' is an invertible O_i, O_k-ideal with $\mathrm{nrd}(J_r') = \mathfrak{p}^r$ and J' is an invertible O_k, O_j-ideal with $\mathrm{nrd}(J') = \mathfrak{p}$. The issue: these products may not all be distinct when they are imprimitive. If the product $J_r'J'$ is imprimitive, then we rewrite it as a compatible product

$$J_r'J' = (J_r'(\overline{J}')^{-1})\overline{J'}J' = \mathfrak{p}(J_r'(\overline{J}')^{-1})$$

where now $J'_{r-1} = \mathfrak{p}^{-1} J'_r J'$ has reduced norm $\mathfrak{p}^{r-1}$. This procedure works in reverse as well.

With apologies for the temporarily annoying notation, define $T_{\text{prim}}(\mathfrak{p}^{r+1})$ and $T_{\text{imprim}}(\mathfrak{p}^{r+1})$ to be the $\mathfrak{p}^r$-Brandt matrix counting classes of primitive or imprimitive, accordingly. Then

$$T(\mathfrak{p}^{r+1}) = T_{\text{prim}}(\mathfrak{p}^{r+1}) + T_{\text{imprim}}(\mathfrak{p}^{r+1}). \tag{41.3.9}$$

Under multiplication by $\mathfrak{p}$, we have

$$T_{\text{imprim}}(\mathfrak{p}^{r+1}) = T(\mathfrak{p}^{r-1})P(\mathfrak{p}). \tag{41.3.10}$$

Since there are $\mathsf{N}(\mathfrak{p})+1$ right O-ideals of reduced norm $\mathfrak{p}$, with the previous paragraph we obtain

$$\begin{aligned} T(\mathfrak{p}^r)T(\mathfrak{p}) &= T_{\text{prim}}(\mathfrak{p}^{r+1}) + (\mathsf{N}(\mathfrak{p})+1)T_{\text{imprim}}(\mathfrak{p}^{r+1}) \\ &= T(\mathfrak{p}^{r+1}) + \mathsf{N}(\mathfrak{p})T_{\text{imprim}}(\mathfrak{p}^{r+1}) \\ &= T(\mathfrak{p}^{r+1}) + \mathsf{N}(\mathfrak{p})T(\mathfrak{p}^{r-1})P(\mathfrak{p}). \end{aligned} \tag{41.3.11}$$

This proves the result for $s=1$, and it gives (41.3.8) upon rearrangement and shifting indices.

We now proceed by (an ugly but harmless) induction on s:

$$\begin{aligned} T(\mathfrak{p}^r)&\left(T(\mathfrak{p}^{s+1}) + \mathsf{N}(\mathfrak{p})T(\mathfrak{p}^{s-1})P(\mathfrak{p})\right) \\ &= \sum_{i=0}^{s}\Big(\mathsf{N}(\mathfrak{p})^i T(\mathfrak{p}^{r+s+1-2i})P(\mathfrak{p})^i \\ &\qquad + \mathsf{N}(\mathfrak{p})^{i+1}P(\mathfrak{p}^{r+s+1-2(i+1)})P(\mathfrak{p})^{i+1}\Big) \end{aligned} \tag{41.3.12}$$

so

$$\begin{aligned} T(\mathfrak{p}^r)T(\mathfrak{p}^{s+1}) &= \sum_{i=0}^{s}\Big(\mathsf{N}(\mathfrak{p})^i T(\mathfrak{p}^{r+s+1-2i}P(\mathfrak{p})^i) \\ &\qquad + \mathsf{N}(\mathfrak{p})^{i+1}P(\mathfrak{p})^{i+1}P(\mathfrak{p}^{r+s+1-2(i+1)})\Big) \\ &\qquad - \sum_{i=0}^{s}\mathsf{N}(\mathfrak{p})^{i+1}P(\mathfrak{p})^{i+1}T(\mathfrak{p}^{r+s+1-2(i+1)}) \\ &= \sum_{i=0}^{s+1}\mathsf{N}(\mathfrak{p})^i P(\mathfrak{p})^i T(\mathfrak{p}^{r+s+1-2i}) \end{aligned} \tag{41.3.13}$$

as claimed. □

Definition 41.3.14. The **Hecke algebra** $\mathbf{T}(O)$ is the subring of $\mathrm{M}_h(\mathbb{Z})$ generated by the matrices $T(\mathfrak{n})$ with $\mathfrak{n}$ coprime to $\mathfrak{N}$.

Corollary 41.3.15. *The ring $\mathbf{T}(O)$ is a commutative $\mathbb{Z}$-algebra.*

Proof. By Proposition 41.3.1(b), we reduce to showing that $T(\mathfrak{p}^r)T(\mathfrak{p}^s) = T(\mathfrak{p}^s)T(\mathfrak{p}^r)$ for all $r, s \geq 0$, and this holds by Proposition 41.3.6: the right-hand side of (41.3.7) is symmetric under interchanging r, s. □

Example 41.3.16. Let $F = \mathbb{Q}(\sqrt{10})$ and $R = \mathbb{Z}_F = \mathbb{Z}[\sqrt{10}]$ its ring of integers. Then the class group $\operatorname{Cl} R \simeq \mathbb{Z}/2\mathbb{Z}$ is nontrivial, represented by the class of the ideal $\mathfrak{p}_2 = (2, \sqrt{10})$, and the narrow class group $\operatorname{Cl}^+ R \simeq \operatorname{Cl} R$ is no bigger: the fundamental unit is $3 + \sqrt{10}$ of norm -1.

Let $B = (-1, -1 \mid F)$. Since 2 is not split in F, the ramification set $\operatorname{Ram} B$ is the set of real places of F. A maximal order is given by

$$O = R \oplus \mathfrak{p}_2^{-1}(1+i) \oplus \mathfrak{p}_2^{-1}(1+j) \oplus R\frac{1+i+j+ij}{2}.$$

We find that $\# \operatorname{Cls} O = 4$, and

$$T(\mathfrak{p}_2) = \begin{pmatrix} 0 & 0 & 0 & 1 \\ 0 & 0 & 3 & 2 \\ 0 & 2 & 0 & 0 \\ 3 & 1 & 0 & 0 \end{pmatrix}.$$

In this case, the matrix $P(\mathfrak{p}_2)$ is the identity matrix: for example, we have $\mathfrak{p}_2 O = (1+i)O$. Thus

$$T(\mathfrak{p}_2^2) = T(2R) = T(\mathfrak{p}_2)^2 - 2 = \begin{pmatrix} 1 & 1 & 0 & 0 \\ 6 & 6 & 0 & 0 \\ 0 & 0 & 4 & 4 \\ 0 & 0 & 3 & 3 \end{pmatrix}.$$

41.4 Semisimplicity

We now equip the space $M_2(O) = \operatorname{Map}(\operatorname{Cls} O, \mathbb{Z})$ with a natural inner product, and we show that the Hecke operators are normal with respect to this inner product.

41.4.1. For $[I] \in \operatorname{Cls} O$, we define $w_{[I]} := [O_{\mathsf{L}}(I)^\times : R^\times]$; this is well-defined, as a different choice of representative gives an isomorphic (conjugate) order.

Let $1_{[I]}$ be the characteristic function of $[I] \in \operatorname{Cls} O$; then $1_{[I]}$ for $[I] \in \operatorname{Cls} O$ form a basis for $M_2(O)$. We define the bilinear form

$$\begin{aligned} \langle\ ,\ \rangle \colon M_2(O) \times M_2(O) &\to \mathbb{Z} \\ \langle 1_{[I]}, 1_{[J]} \rangle &:= w_{[I]} \delta_{[I],[J]} \end{aligned} \tag{41.4.2}$$

where $\delta_{[I],[J]} = 1, 0$ according as $[I] = [J]$ or not, and extend linearly. The matrix of this pairing in the basis of characteristic functions is the diagonal matrix $\operatorname{diag}(w_i)_i$, where $w_i = [O_i^\times : R^\times]$. The pairing is symmetric and nondegenerate.

The inner product (41.4.2) defines an adjoint $T \mapsto T^*$.

Proposition 41.4.3. *We have*

$$P(\mathfrak{n})^* = P(\mathfrak{n}^{-1}) \tag{41.4.4}$$
$$T(\mathfrak{n})^* = P(\mathfrak{n}^{-1})T(\mathfrak{n}). \tag{41.4.5}$$

The Hecke operators $T(\mathfrak{n})$ *are normal with respect to the inner product (41.4.2), and for* $\mathfrak{n}$ *trivial in* $\mathrm{Cl}^+ R$ *the operators* $T(\mathfrak{n})$ *are self-adjoint.*

Proof. We may show the proposition for the Brandt matrices. Let $W = \mathrm{diag}(w_i)_i$ define the inner product on $\mathbb{Z}^h$ with the Brandt matrices acting on the *right* on row vectors. Then the inner product is $\langle x, y\rangle = xWy^{\mathsf{t}}$ and accordingly the adjoint $\langle xT, y\rangle = \langle x, T^* y\rangle$ is defined by

$$T^* = W^{-1}T^{\mathsf{t}}W \tag{41.4.6}$$

The transpose of a permutation matrix is its inverse and that $O_{\mathsf{L}}(\mathfrak{n}I_i) = O_{\mathsf{L}}(I_i)$, so that the unit groups match up, whence

$$P(\mathfrak{n})^* = P(\mathfrak{n})^{-1} = P(\mathfrak{n}^{-1}) \tag{41.4.7}$$

giving (41.4.4).

For the Brandt matrices, we refer to Lemma 41.2.7(a), giving

$$T(\mathfrak{n})_{ij} = \frac{1}{w_i}\#\left\{\alpha \in I_j I_i^{-1} : \mathrm{nrd}(\alpha)R = \mathfrak{n}_{ij}\right\}/R^\times$$

where $\mathfrak{n}_{ij} = \mathfrak{n}\,\mathrm{nrd}(I_j)/\,\mathrm{nrd}(I_i)$. Let

$$\Theta(\mathfrak{n})_{ij} = \left\{\alpha \in I_j I_i^{-1} : \mathrm{nrd}(\alpha)R = \mathfrak{n}_{ij}\right\}/R^\times.$$

By (41.4.6),

$$WT(\mathfrak{n}) = (\Theta(\mathfrak{n})_{ij})_{i,j} =: \Theta(\mathfrak{n}).$$

We extend the definition of $\Theta(\mathfrak{n})$ to include all fractional ideals $\mathfrak{n}$. For each i, let i' be such that $[\mathfrak{n}^{-1}I_i] = [I_{i'}]$, so that $\mathfrak{n}^{-1}I_i = \beta_i I_i'$; the induced action is given by the permutation map $P(\mathfrak{n}^{-1})$.

$$\begin{aligned}\Theta(\mathfrak{n})_{ij} &\to \Theta(\mathfrak{n})_{ji'}\\ \alpha &\mapsto (\alpha\beta_i)^{-1} = \beta_i^{-1}\alpha^{-1}\end{aligned} \tag{41.4.8}$$

is well-defined and bijective.

Indeed, if $\alpha \in I_j I_i^{-1}$ then

$$\overline{\alpha} \in \overline{I_j I_i^{-1}} = \overline{I_i^{-1}}\,\overline{I_j} = I_i I_j^{-1}\frac{\mathrm{nrd}(I_j)}{\mathrm{nrd}(I_i)} \tag{41.4.9}$$

since $I\overline{I} = \mathrm{nrd}(I)$ for an invertible R-lattice I, so

$$\alpha^{-1} \in \mathfrak{n}^{-1}I_i I_j^{-1} = \beta_i I_{i'} I_j^{-1}$$

and therefore $\beta_i^{-1}\alpha^{-1} \in I_{i'}I_j^{-1}$ as claimed. And $\operatorname{nrd}(\alpha) = \mathfrak{n}_{ij}$ implies $\operatorname{nrd}(\alpha^{-1}) = \mathfrak{n}^{-2}\mathfrak{n}_{ji}$ so $\operatorname{nrd}(\beta_i^{-1}\alpha^{-1}) = \mathfrak{n}_{ji'}$. We can run the argument in the other direction to produce an inverse, and we thereby conclude the map is bijective.

The map (41.4.8) together with the action by permutation and $W^{\mathsf{t}} = W$ yields

$$WT(\mathfrak{n}) = \Theta(\mathfrak{n}) = P(\mathfrak{n}^{-1})\Theta(\mathfrak{n})^{\mathsf{t}} = P(\mathfrak{n}^{-1})WT(\mathfrak{n})^*$$

and thus $T(\mathfrak{n})^* = P(\mathfrak{n})^*T(\mathfrak{n})$, and substituting (41.4.7) gives (41.4.5).

For the final statement, by (41.3.5) we have $T(\mathfrak{n})$ commuting with $P(\mathfrak{n})$, so $T(\mathfrak{n})$ commutes with $T(\mathfrak{n})^*$; and when $\mathfrak{n}$ is narrowly principal, then $P(\mathfrak{n}^{-1})$ is the identity matrix so $T(\mathfrak{n})^* = T(\mathfrak{n})$. □

By the spectral theorem in linear algebra, we have the following corollary.

Corollary 41.4.10. *$\boldsymbol{T}(O)$ is a semisimple commutative ring, and there exists a basis of common eigenvectors (eigenfunctions) for the Hecke operators. Each $T(\mathfrak{n})$ with $\mathfrak{n}$ narrowly principal has real eigenvalues.*

41.5 Eichler trace formula

In this section, we compute the trace of the Brandt matrices in terms of embedding numbers. We continue notation from the previous section.

We begin by recalling the main ingredients. Let $K \supset F$ be a separable quadratic field extension and let $S \subseteq K$ be a quadratic R-order. Let $h(S) = \#\operatorname{Pic} S$. Let $m(S, O, O^\times)$ be the number of $O^\times$-conjugacy classes of optimal embeddings $S \hookrightarrow O$. Then by Theorem 30.4.7,

$$\sum_{[I]\in\operatorname{Cls} O} m(S, O_{\mathsf{L}}(I); O_{\mathsf{L}}(I)^\times) = h(S)m(\widehat{S}, \widehat{O}; \widehat{O}^\times) \tag{41.5.1}$$

We also recall

$$\operatorname{mass}(\operatorname{Cls} O) := \sum_{i=1}^{h} \frac{1}{w_i}$$

and that the Eichler mass formula (Main Theorem 26.1.5) gives an explicit formula for this mass in terms of the relevant arithmetic invariants.

Main Theorem 41.5.2 (Trace formula). *If $\mathfrak{n}$ is not narrowly principal, then* $\operatorname{tr} T(\mathfrak{n}) = 0$. *If $\mathfrak{n} = nR$ is narrowly principal with $n \in F_{>0}^\times$, then*

$$\operatorname{tr} T(\mathfrak{n}) = \frac{1}{2}\sum_{(u,t,S)} \frac{h(S)}{w_S} m(\widehat{S}, \widehat{O}; \widehat{O}^\times) + \begin{cases} \operatorname{mass}(\operatorname{Cls} O), & \text{if } \mathfrak{n} = c^2R,\ c \in F^\times \\ 0, & \text{otherwise} \end{cases}$$

where $w_S := [S^\times : R^\times]$ and the sum is over finitely many triples (u, t, S) where:

- *u belongs to a set of representatives of $R_{>0}^\times / R^{\times 2}$;*
- *$t \in R$ satisfies $t^2 - 4un \in F_{<0}^\times$; and*
- *$S \supseteq R[x]/(x^2 - tx + un)$.*

Proof. We have $\operatorname{tr} T(\mathfrak{n}) = \sum_{i=1}^k T(\mathfrak{n})_{ii}$. By Lemma 41.2.7, since $\mathfrak{n}_{ii} = \mathfrak{n}$ we conclude $\operatorname{tr} T(\mathfrak{n}) = 0$ if $\mathfrak{n}$ is not narrowly principal. So suppose $\mathfrak{n} = nR$ is narrowly principal, with $n \in F_{>0}^\times$. Then by (41.2.9) we have

$$w_i T(\mathfrak{n})_{ii} = \frac{1}{2} \sum_{uR^{\times 2} \in R_{>0}^\times / R^{\times 2}} \#\{\alpha \in \mathcal{O}_i : \operatorname{nrd}(\alpha) = un\}. \tag{41.5.3}$$

We are free to organize by reduced trace, giving

$$w_i T(\mathfrak{n})_{ii} = \frac{1}{2} \sum_u \sum_{t \in R} \#\{\alpha \in \mathcal{O}_i : \operatorname{trd}(\alpha) = t,\ \operatorname{nrd}(\alpha) = un\}. \tag{41.5.4}$$

Since B is definite, we have $\operatorname{disc}(\alpha) = t^2 - 4un$ either zero or totally negative, so the inner sum is over finitely many $t \in R$ either satisfying $t^2 = 4un$ or $t^2 - 4un \in F_{<0}^\times$.

If $t^2 = 4un$ (equivalently $\alpha = t/2 \in F$), then $\mathfrak{n} = nR = c^2 R$ with $c = \pm t/2$; conversely, if $\mathfrak{n} = nR = c^2 R$ for some $c \in F^\times$, then there exists a unique representative $u \in R_{>0}^\times / R^{\times 2}$ such that $un = c^2$. Consequently, exactly when $\mathfrak{n} = c^2 R$ is a square of a principal ideal, there is a contribution of $(1/2)(2) = 1$ to the sum.

For the remaining terms, we have $\alpha \notin F$ and $R[\alpha] \simeq R[x]/(x^2 - tx + un)$ is a domain. The embedding $R[\alpha] \hookrightarrow \mathcal{O}_i$ need not be optimal, but nevertheless corresponds to the optimal embedding $S \hookrightarrow \mathcal{O}_i$ for a unique superorder $S \supseteq R[\alpha]$, and conversely. We count these up to units: the action of conjugaton by $\mu \in \mathcal{O}_i^\times$ centralizes such an embedding if and only if $\mu \in S^\times$, so letting $w_S := [S^\times : R^\times]$ we have

$$\#\{\alpha \in \mathcal{O}_i : \operatorname{trd}(\alpha) = t,\ \operatorname{nrd}(\alpha) = un\} = \sum_{S \supseteq R[x]/(x^2 - tx + un)} m(S, \mathcal{O}_i; \mathcal{O}_i^\times) \frac{w_i}{w_S}.$$

Plugging these into (41.5.4), we obtain

$$\begin{aligned} w_i T(\mathfrak{n})_{ii} = \frac{1}{2} \sum_{u,t} \sum_{S \supseteq R[x]/(x^2 - tx + un)} & m(S, \mathcal{O}_i; \mathcal{O}_i^\times) \frac{w_i}{w_S} \\ & + \begin{cases} 1, & \text{if } \mathfrak{n} = (cR)^2; \\ 0, & \text{otherwise.} \end{cases} \end{aligned} \tag{41.5.5}$$

Dividing through by w_i and summing (41.5.5), we have

$$\operatorname{tr} T(\mathfrak{n}) = \frac{1}{2} \sum_{(u,t,S)} \sum_{i=1}^h \frac{1}{w_S} m(S, \mathcal{O}_i; \mathcal{O}_i^\times) + \delta \sum_i \frac{1}{w_i} \tag{41.5.6}$$

where $\delta = 1, 0$ according as $\mathfrak{n}$ is a square of a principal ideal or not. Now substituting (41.5.1) and the definition of mass, the theorem is proven. $\square$

Corollary 41.5.7. *We have*

$$\# \operatorname{Cls} \mathcal{O} = \operatorname{mass}(\operatorname{Cls} \mathcal{O}) + \frac{1}{2} \sum_{(u,t,S)} \frac{h(S)}{w_S} m(\widehat{S}, \widehat{\mathcal{O}}; \widehat{\mathcal{O}}^\times)$$

where the sum is over (u, t, S) with $n = 1$.

Proof. For $\mathfrak{n} = R$, we have $T(1)$ the identity matrix, so $\operatorname{tr} T(1) = \#\operatorname{Cls} O$. □

Corollary 41.5.7 gives a different way to prove (and interpret) the Eichler class number formula (Main Theorem 30.8.6): for the exact comparison, see Exercise 41.2.

Example 41.5.8. Suppose $F = \mathbb{Q}$ and $R = \mathbb{Z}$. For $d \in \mathbb{Z}$ a nonsquare discriminant, we write $S_d := \mathbb{Z}[(d + \sqrt{d})/2]$ for the unique quadratic ring of discriminant d, so $w_{S_d} = [S_d^\times : \mathbb{Z}^\times] = 1$ except for $d = -3, -4$. In the trace formula (Main Theorem 41.5.2), the quaternion algebra that appears is definite, and so the only quadratic orders that embed are imaginary quadratic, with discriminant $d < 0$. Simplifying in this way, the trace formula then becomes

$$\operatorname{tr} T(n) = \frac{1}{2} \sum_{t \in \mathbb{Z}} \sum_{df^2 = t^2 - 4n < 0} \frac{h(S_d)}{w_{S_d}} m(\widehat{S}_d, \widehat{O}; \widehat{O}^\times) \tag{41.5.9}$$

for n not a square (adding a mass term for n a square).

To notationally simplify a bit further, we define *modified* Hurwitz class numbers

$$h_O(S) := \frac{h(S)}{w_S} m(\widehat{S}, \widehat{O}; \widehat{O}^\times)$$

where the factor $m(\widehat{S}, \widehat{O}; \widehat{O}^\times)$ is defined by purely local data, given in section 30.5 for maximal orders and section 30.6 for Eichler orders. Writing $h_O(d) = h_O(S_d)$ for the order of discriminant d, we arrive at a pleasing formula:

$$\operatorname{tr} T(n) = \frac{1}{2} \sum_{t \in \mathbb{Z}} \sum_{df^2 = t^2 - 4n < 0} h_O(d) \tag{41.5.10}$$

again for n not a square.

Taking $n = 1$ as in Corollary 41.5.7 (and adding back the mass term) gives

$$\begin{aligned} \#\operatorname{Cls} O &= \operatorname{mass}(\operatorname{Cls} O) + \frac{1}{2} \sum_{\substack{t \in \mathbb{Z} \\ t^2 < 4}} h_O(t^2 - 4) \\ &= \operatorname{mass}(\operatorname{Cls} O) + \frac{1}{2} h_O(-4) + 2 h_O(-3). \end{aligned} \tag{41.5.11}$$

For O an Eichler order, after substitution we recover the Eichler class number formula (Theorem 30.1.5).

Example 41.5.12. As an illustration of Example 41.5.8, we compute $\operatorname{tr} T(3)$ for the Hurwitz order. We compute the following table of values:

t	d	$h(S_d)$	w_{S_d}	$m(\widehat{S}_d, \widehat{O}; \widehat{O}^\times)$	$h_O(d)$
0	−12	1	1	0	0
0	−3	1	3	2	2/3
±1	−11	1	1	$1 - (-11 \mid 2) = 2$	2
±2	−8	1	1	$1 - (-8 \mid 2) = 1$	1
±3	−3	1	3	$1 - (-3 \mid 2) = 2$	2/3

Summing then gives

$$\operatorname{tr} T(3) = (1/2)(0 + 2/3) + (2 + 1 + 2/3) = 4.$$

Indeed, more generally since the Hurwitz order has $\#\operatorname{Cls} O = 1$, the matrix $T(n)$ is a 1×1-matrix with $T(n) = [\sigma(n)]$ for n odd. This observation implies a nontrivial (and otherwise surprising) relationship between class numbers of imaginary quadratic orders!

Remark 41.5.13. Brandt [Bra43, §III] defined Brandt matrices in the same paper as his groupoid; he called them *Hecke matrices*, as he claimed to follow parallels with certain operators defined by Hecke. Indeed, Hecke [Hec40, §9, Satz 53] conjectured that the space of cusp forms of weight 2 on $\Gamma_0(p)$ for p prime was spanned by certain linear combinations of theta series, and it was this observation that motivated Brandt. (Eichler [Eic56a, footnote 16] says that Brandt should not have named them after Hecke, since it was really Brandt who interpreted function-theoretic results of Hecke using pure arithmetic.)

Eichler [Eic56a] proved that the ring generated by the Brandt matrices was a commutative, semisimple ring and proved the trace formula for Brandt matrices [Eic56a, §6]. In this early work, he already foresaw the application of Brandt matrices to other base fields: as an application, he used Brandt matrices to give class number relations between imaginary quadratic fields, and in the function field case these become relations among divisor class groups for hyperelliptic curves. Eichler [Eic77, Chapter II] presented the generalization to totally real fields, giving a treatment of Hecke operators, Brandt matrices, and theta series, and he proved that the Brandt matrices realize Hecke operators in certain spaces of Hilbert modular forms.

Eichler then later gave a self-contained presentation [Eic73, Chapter II] of the theory of Brandt matrices over $\mathbb{Q}$, with the intended application the solution to Hecke's conjecture (suitably corrected), now known as the **basis problem** for $\Gamma_0(p)$: to give bases of linearly independent forms of spaces of modular forms in terms of theta series of quadratic forms coming from quaternion algebras. This line of work was followed by generalizations by Hijikata [Hij74] and Hijikata–Saito [HS73] for general Eichler orders, Pizer [Piz76b, Piz76c] for residually split orders, culminating in a solution over the rational numbers to the basis problem by Hijikata–Pizer–Shemanske [HPS89a].

The method of proof for the solution to the basis problem is the use of the trace formula, for which a key ingredient is the theory of optimal embeddings: see Remark 30.6.18.

Indeed, it is much more involved analytically, but one can similarly compute the trace of the Hecke operator acting on classical spaces of modular forms or more generally spaces of Hilbert modular forms. These trace formulae are quite complicated, but one notices that they have a similar shape as the above trace formula; and in fact, under certain hypotheses and after restricting to an appropriate new subspace, the traces are *equal*. But since both rings are semisimple, this implies that the same systems of eigenvalues for the Hecke operators arise! Such a correspondence was first given by Eichler, as above; it was generalized to totally real fields by Shimizu [Shz72] using theta series, and the most general formulation given by Jacquet–Langlands [JL70]. This correspondence was conjectured to generalize to the principle of *Langlands functorial transfer*: for an introduction to this vast area, see Gelbart [Gel84].

In light of the preceding epic remark, we hope we have inspired the reader to pursue the relationship between Brandt matrices and modular forms! Unfortunately, it would require another book to respectfully develop this subject.

Remark 41.5.14. Pizer [Piz80a] was the first to give an algorithm for computing classical modular forms using Brandt matrices (on $\Gamma_0(N)$ for N not a perfect square); see also the work of Kohel [Koh2001] over $\mathbb{Z}$. This algorithm was generalized to compute Hilbert modular forms over a totally real field of narrow class number 1 by Socrates–Whitehouse [SW2005], with algorithmic improvements by Dembélé [Dem2007]. The assumption on the class number was removed by Dembélé–Donnelly [DD2008]. A survey of these methods are given by Dembélé–Voight [DV2013, §4, §8].

Exercises

Unless otherwise specified, in these exercises let R be a global ring with eligible set $S \subseteq \operatorname{Pl} F$, let B be an S-definite quaternion algebra over F and let $O \subset B$ be an R-order in B.

1. Extend the definition of the Brandt matrix to include the case $\mathfrak{n} = (0)$ of the zero ideal, following (41.2.8): define $T(0)_{ij} = 1/w_i$ for $i, j = 1, \dots, h$. Conclude $\operatorname{tr} T(0) = \operatorname{mass}(\operatorname{Cls} O)$.
2. Show that Corollary 41.5.7 agrees with Main Theorem 30.8.6. *[Hint: organize by $q := [S^\times : R^\times]$, observe that in $S \supseteq R[x]/(x^2 - tx + u)$ we have necessarily $[S^\times : R^\times] \geq 2$ and each such S contains $q-1$ orders of the form $R[x]/(x^2-tx+u)$.]*
3. Refine Lemma 41.2.7(c) in a special case as follows. Suppose $\mathrm{Cl}^+ R$ is trivial. Show that
$$T(\mathfrak{n})_{ij} = \frac{1}{w_{i,1}} \#\{\alpha \in I_j I_i^{-1} : \operatorname{nrd}(\alpha) = n_{ij}\}$$
where $w_{i,1} := \#O_i^1$.
4. Suppose $R = \mathbb{Z}$ and $F = \mathbb{Q}$, suppose $\operatorname{disc} B = p$ is prime and O is a maximal order. Show that
$$\operatorname{tr} T(p) = \begin{cases} 1, & \text{if } p = 2, 3; \\ h_O(-4p), & \text{if } p > 3. \end{cases}$$
where
$$h_O(-4p) = \begin{cases} h(-4p)/2, & \text{if } p \equiv 1 \pmod 4; \\ h(-p), & \text{if } p \equiv 7 \pmod 8; \\ 2h(-p), & \text{if } p \equiv 3 \pmod 8 \text{ and } p > 3. \end{cases}$$
What does this say about the number of maximal orders in B up to isomorphism such that every *two-sided* ideal is principal?
5. Give another proof of Proposition 41.3.1(b) using the local-global dictionary for lattices.
6. Prove (41.3.12) using induction and then expand to verify (41.3.13).
7. Let $B = \left(\dfrac{-1, -11}{\mathbb{Q}}\right)$ with $\operatorname{disc} B = 11$ and let $O = \mathbb{Z}\langle i, \tfrac{1}{2}(j+1)\rangle$.

(a) Show that O is a maximal order with $\#O^\times = 4$.
(b) Show that the ternary quadratic form associated to O is similar to $x^2 - xz + y^2 + 3z^2$.
(c) Show that $\operatorname{Cl} O = \{[O], [I_2]\}$ where $I_2 = 2O + \frac{1}{2}(1+2i+j)O$. *[Hint: Follow Example* 17.6.3.] Along the way, show that

$$T(2) = \begin{pmatrix} 1 & 3 \\ 2 & 0 \end{pmatrix}.$$

(d) Pause and show that $O_2 := O_{\mathsf{L}}(I_2)$ has $\#O_2^\times = 6$ and associated ternary quadratic form $x^2 - xy - xz + y^2 + yz + 4z^2$.
(e) Show that $M_2(O)$ has two eigenspaces for the Hecke algebra, one spanned by a form e with $T(p)(e) = (p+1)e$ for all $p \neq 11$, and the other spanned by a form f with $T(2)(f) = -2f$.
(f) Verify the trace formula (41.5.9) for $\operatorname{tr} T(2) = 1$ by computing class numbers.

[There is a unique normalized cusp form $f \in S_2(\Gamma_0(11))$ of weight 2 and level 11 with

$$f(q) = q \prod_{n=1}^{\infty} (1-q^n)^2 \prod_{n=1}^{\infty} (1-q^{11n})^2 = q - 2q^2 - q^3 + \cdots = \sum_{n=1}^{\infty} a_n q^n$$

matching f in the sense that $T(n)f = a_n f$ for all $11 \nmid n$.]

8. Let $O(3,5)$ be an Eichler order of level 5 and reduced discriminant 15 (in a quaternion algebra B of discriminant 5) and similarly let $O(5,3)$ be an Eichler order of level 3 and reduced discriminant 15.

(a) Show that $\#\operatorname{Cls} O(3,5) = \#\operatorname{Cls} O(5,3) = 2$.
(b) Show that there is a unique eigenvector for the Brandt matrix $T(2)$ for both orders with eigenvalue -1. As far as you can compute, show that this eigenvector shares the same eigenvalues for $T(n)$ both orders.

9. We consider an example of Brandt matrices not restricted to maximal orders. Let $B = \left(\dfrac{-1,-1}{\mathbb{Q}}\right)$ and let

$$O = \mathbb{Z}\langle 2i, 2j\rangle = \mathbb{Z} \oplus \mathbb{Z}(2i) \oplus \mathbb{Z}(2j) \oplus \mathbb{Z}(4ij).$$

(a) Show that O is an order with $\operatorname{discrd} O = N = 64$.
(b) Compute that $\#\operatorname{Cls} O = 4$.
(c) Under the action of the Brandt matrices $T(n)$, show that there are 3 irreducible factors of dimensions 1, 1, 2. In a basis of characteristic functions, identify the one-dimensional factors as:

$$(1,1,1,1) \leftrightarrow e(q) := \frac{1}{24} + \sum_{n=1}^{\infty} \sigma^*(n) q^n$$

$$(1,-1,-1,1) \leftrightarrow e_\chi(q) := \frac{1}{24} + \sum_{n=1}^{\infty} \sigma^*(n)\chi(n) q^n$$

where now $\sigma^*(n) := \sum_{2\nmid d|n} d$ and $\chi(n) = \left(\frac{-1}{n}\right)$.

[The two-dimensional space has basis $(1, 0, 0, -1), (0, 1, -1, 0) \leftrightarrow f_1, f_2$ and $a_p(f_1) = a_p(f_2)$ for all p, with

$$f_i = q + 2q^5 - 3q^9 - 6q^{13} + 2q^{17} + \dots$$

corresponding to the isogeny class of the elliptic curve $E\colon y^2 = x^3 + x$ of conductor 64.]

This chapter is licensed under the terms of the Creative Commons Attribution-NonCommercial 4.0 International License (http://creativecommons.org/licenses/by-nc/4.0/), which permits any noncommercial use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license and indicate if changes were made.

The images or other third party material in this chapter are included in the chapter's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the chapter's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.

Chapter 42

Supersingular elliptic curves

In the previous chapter, we showed that Brandt matrices for an order in a definite quaternion algebra B contain a wealth of arithmetic. In the special case where $\operatorname{disc} B = p$ is prime, there is a further beautiful connection between Brandt matrices and the theory of supersingular elliptic curves, arising from the following important result: there is an equivalence of categories between supersingular elliptic curves over $\overline{\mathbb{F}}_p$ and right ideals in a (fixed) maximal order $O \subset B$. We pursue this important connection in this chapter for the reader who has a bit more background in algebraic curves.

42.1 Supersingular elliptic curves

In this section, we briefly review what we will need from the theory of elliptic curves; see Silverman [Sil2009] for further general reference. Let F be a field with algebraic closure F^{al}.

Definition 42.1.1. An **elliptic curve** is a smooth projective **curve** (variety of dimension 1) of genus 1 equipped with a rational point. Every elliptic curve E is isomorphic over F to the projective curve associated to the affine equation

$$E : y^2 + a_1xy + a_3y = x^3 + a_2x^2 + a_4x + a_6$$

with $a_i \in F$.

Definition 42.1.2. An **isogeny** $\phi\colon E \to E'$ is a nonconstant morphism of pointed curves; such a map is automatically surjective and a group homomorphism, with the marked point as origin.

Let $\operatorname{Hom}(E, E')$ be the collection of isogenies from E to E' defined over F; if we need to allow isogenies defined over a larger field, we will similarly extend the field of definition of our elliptic curves. Then $\operatorname{Hom}(E, E')$ is a torsion-free $\mathbb{Z}$-module of rank at most four. Let $\operatorname{End}(E) := \operatorname{Hom}(E, E)$ be the **endomorphism ring** of E and let $\operatorname{End}(E)_{\mathbb{Q}} := \operatorname{End}(E) \otimes_{\mathbb{Z}} \mathbb{Q}$ be the **endomorphism algebra**.

42.1.3. For each nonzero isogeny $\phi\colon E \to E'$, there exists a **dual isogeny** $\phi^\vee : E' \to E$ such that $\phi^\vee \circ \phi$ and $\phi \circ \phi^\vee$ are equal to multiplication by the degree $\deg \phi \in \mathbb{Z}_{>0}$ on

© The Author(s) 2021

J. Voight, *Quaternion Algebras*, Graduate Texts in Mathematics 288,
https://doi.org/10.1007/978-3-030-56694-4_42

E and E', respectively. In particular, the dual $^\vee$ is a standard involution on $\mathrm{End}(E)$ that is positive (see 8.4.1); the $\mathbb{Q}$-algebra $\mathrm{End}(E)_\mathbb{Q} := \mathrm{End}(E) \otimes_\mathbb{Z} \mathbb{Q}$ is therefore a division ring. In particular, we have the equality

$$\phi^{-1} = \frac{1}{\deg \phi}\phi^\vee \tag{42.1.4}$$

in $\mathrm{End}(E)_\mathbb{Q}$ for all nonzero $\phi \in \mathrm{End}(E)$.

From now on, let E be an elliptic curve over F.

Lemma 42.1.5. *The endomorphism algebra* $\mathrm{End}(E)_\mathbb{Q}$ *of E is either $\mathbb{Q}$, an imaginary quadratic field K, or a definite quaternion algebra over $\mathbb{Q}$.*

Proof. We apply Theorem 3.5.1 to conclude that $\mathrm{End}(E)_\mathbb{Q}$ is either $\mathbb{Q}$, a quadratic field, or a division quaternion algebra. Then by Example 8.4.2, the involution is positive if and only if $\mathrm{End}(E)_\mathbb{R}$ is $\mathbb{R}$, $\mathbb{C}$, or $\mathbb{H}$, so in the second case we must have an *imaginary* quadratic field and in the third case we must have a *definite* quaternion algebra. □

42.1.6. Among the possibilities in Lemma 42.1.5, if $\mathrm{End}(E_{F^{\mathrm{al}}})_\mathbb{Q}$ is a quaternion algebra, then we say E is **supersingular**.

See Silverman [Sil2009, §V.3] for a treatment of supersingular elliptic curves.

Proposition 42.1.7. *If E is supersingular, then* $\mathrm{char}\, F = p > 0$. *Moreover, the following are equivalent:*

(i) *E is supersingular;*
(ii) $E[p](F^{\mathrm{al}}) = \{0\}$*; and*
(iii) *the map $[p] : E \to E$ is purely inseparable and $j(E) \in \mathbb{F}_{p^2}$;*

If F is a finite field, then these are further equivalent to

(iv) $\mathrm{trd}(\phi) = \phi + \phi^\vee \equiv 0 \pmod p$*, where $\phi\colon E \to E$ is the Frobenius endomorphism.*

Proof. See Silverman [Sil2009, V.3.1]. □

42.1.8. One can often reduce questions about supersingular elliptic curves to ones where the base field F is $\mathbb{F}_{p^2}$ as follows: by Proposition 42.1.7(iii), if E is supersingular then E is isomorphic over F^{al} to a curve E defined over $\mathbb{F}_{p^2}$.

The following fundamental result is due to Deuring [Deu41]; we give a proof due to Lenstra [Len96, §3].

Theorem 42.1.9. *Let E be an elliptic curve over F and suppose that* $\mathrm{rk}_\mathbb{Z} \mathrm{End}(E) = 4$. *Then* $B = \mathrm{End}(E)_\mathbb{Q}$ *is a quaternion algebra over $\mathbb{Q}$ ramified at* $p = \mathrm{char}\, F$ *and ∞, and* $\mathrm{End}(E)$ *is a maximal order in B.*

In particular, if over F we have $\dim \mathrm{End}(E) = 4$, then automatically E has all of its endomorphisms defined over F.

Proof. Let $O = \operatorname{End} E \subseteq B = \operatorname{End}(E)_{\mathbb{Q}}$. Let $n > 0$ be prime to p. Then there is an isomorphism [Sil2009, Corollary III.6.4(b)]

$$E[n] = E[n](F^{\mathrm{al}}) \simeq \mathbb{Z}/n\mathbb{Z} \oplus \mathbb{Z}/n\mathbb{Z}$$

as abelian groups, and the endomorphism ring of this abelian group is $\operatorname{End} E[n] \simeq \mathrm{M}_2(\mathbb{Z}/n\mathbb{Z})$.

We claim that the structure map $O/nO \to \operatorname{End} E[n]$ is injective, which is to say, $E[n]$ is a faithful module over O/nO. Indeed, suppose $\phi \in O$ annihilates $E[n]$; then since multiplication by n is separable, by the homomorphism theorem for elliptic curves [Sil2009, Corollary III.4.11] there exists $\psi \in O$ such that $\phi = n\psi$, so $\phi \equiv 0 \in O/nO$, proving injectivity. But further, since $\#O/nO = \#\operatorname{End} E[n] = n^4$, the structure map is an isomorphism.

Since O is a free $\mathbb{Z}$-module, we have

$$O_\ell := O \otimes_{\mathbb{Z}} \mathbb{Z}_\ell = O \otimes_{\mathbb{Z}} \varprojlim_n \mathbb{Z}/\ell^n\mathbb{Z} \simeq \varprojlim_n O/\ell^n O.$$

The structure isomorphisms in the previous paragraph are compatible with respect to powers of ℓ, so with the previous line they provide an isomorphism

$$O_\ell \xrightarrow{\sim} \varprojlim_n \operatorname{End} E[\ell^n] = \operatorname{End}_{\mathbb{Z}_\ell} E[\ell^\infty] \simeq \mathrm{M}_2(\mathbb{Z}_\ell)$$

of $\mathbb{Z}_\ell$-algebras, and in particular O_ℓ is maximal and $B_\ell \simeq \mathrm{M}_2(\mathbb{Q}_\ell)$ so B is split at ℓ.

Since B is definite, it follows from the classification theorem (Main Theorem 14.1.3, equivalent to quadratic reciprocity) that $\operatorname{Ram}(B) = \{p, \infty\}$, so B_p is a division algebra over $\mathbb{Q}_p$.

To conclude, we show that O_p is maximal. For $\phi \in O$ an isogeny, let $\deg_i \phi$ be the inseparable degree of ϕ, which is a power of p. We put $\deg_i 0 = \infty$. Then $\deg_i \phi$ is divisible by $q = p^r$ if and only if ϕ factors via the qth power Frobenius morphism $E \to E^{(q)}$. The map

$$\begin{aligned} v : \operatorname{End}(E)_{\mathbb{Q}} &\to \mathbb{Q} \cup \{\infty\} \\ v(a\phi) &= \operatorname{ord}_p(a) + \frac{1}{2}\operatorname{ord}_p(\deg_i \phi) \end{aligned} \tag{42.1.10}$$

for $a \in \mathbb{Q}$ and $\phi \in \operatorname{End}(E)$ is well-defined (since $\deg_i[p] = \deg[p] = p^2$). Factoring an isogeny into its separable and inseparable parts shows that

$$\operatorname{ord}_p(\deg_i \phi) = \operatorname{ord}_p(\deg \phi) = \operatorname{ord}_p(\operatorname{nrd} \phi)$$

so (42.1.10) is precisely the valuation (13.3.1) on $B = \operatorname{End}(E)_{\mathbb{Q}}$ extending the p-adic valuation on $\mathbb{Q}$. (See also Exercise 42.2.)

To conclude, we show that $O_{(p)}$ is the valuation ring (13.3.3) of B and is therefore maximal (Proposition 13.3.4). If $\alpha \in O_{(p)} = O \otimes_{\mathbb{Z}} \mathbb{Z}_{(p)}$ then $\deg \alpha \in \mathbb{Z}_{(p)}$ so α is in the valuation ring. Conversely, let $\alpha \in B$ be a rational isogeny with $v(\alpha) \geq 0$, and write $\alpha = a\phi$ where ϕ is an (actual) isogeny not divisible by any integer. Then

$\nu(\alpha) = \operatorname{ord}_p(a) + \nu(\phi) \geq 0$ and $0 \leq \nu(\phi) \leq 1/2$, since the multiplication by p is purely inseparable; so $\operatorname{ord}_p(a) \geq -1/2$ and therefore $a \in \mathbb{Z}_{(p)}$, and hence $\alpha \in O_{(p)}$.

Finally, since an order is maximal if and only if it is locally maximal, O itself is a maximal order in the quaternion algebra B. □

In light of 42.1.8, we now let $F = \mathbb{F}_p^{\text{al}}$ be an algebraic closure of $\mathbb{F}_p$. Let E, E' be elliptic curves over F. If E is isogenous to E', then E is supersingular if and only if E' is supersingular (see Exercise 42.1). The converse is also true, as follows.

Lemma 42.1.11. *Let E, E' be supersingular elliptic curves over F. Then* $\operatorname{Hom}(E, E')$ *is a $\mathbb{Z}$-module of rank 4 that is invertible as a right* $\operatorname{End}(E)$*-module under precomposition and a left* $\operatorname{End}(E')$*-module under postcomposition.*

In particular, if E, E' are supersingular elliptic curves over $\mathbb{F}_p^{\text{al}}$, then there exists a separable isogeny $E \to E'$.

Proof. We may suppose E is defined over a finite field $\mathbb{F}_q$ such that E has all of its endomorphisms defined over $\mathbb{F}_q$. Let $\pi \in O = \operatorname{End}(E)$ be the q-power Frobenius endomorphism. Then $B = O \otimes_{\mathbb{Z}} \mathbb{Q}$ is a quaternion algebra over $\mathbb{Q}$. Since $\operatorname{End}(E)$ is defined over $\mathbb{F}_q$, the endomorphism π commutes with every isogeny $\alpha \in O$, and so π lies in the center of O; since $Z(B) = \mathbb{Q}$, we have $\pi \in \mathbb{Z} = Z(O)$. But $\deg \pi = \pi\overline{\pi} = \pi^2 = q$ so $\pi = \pm\sqrt{q} \in \mathbb{Z}$. Therefore $\#E(\mathbb{F}_q) = q + 1 \mp 2\sqrt{q}$. Therefore $\#E(\mathbb{F}_{q^2}) = q^2 + 1 - 2q = (q-1)^2$.

Continuing to enlarge $\mathbb{F}_q$, we may repeat the above argument with E' to conclude that $\#E(\mathbb{F}_q) = \#E'(\mathbb{F}_q)$. It then follows that E, E' are isogenous over $\mathbb{F}_q$ [Sil2009, Exercise III.5.4(b)], but we will show this and more. Let $\ell \neq p$ be prime and let $T_\ell(E) = \varprojlim_n E[\ell^n] \simeq \mathbb{Z}_\ell^2$ is the ℓ-adic Tate module of E. By the Isogeny Theorem [Sil2009, Theorem III.7.7(a)], for every prime $\ell \neq p$, the natural map

$$\operatorname{Hom}_{\mathbb{F}_q}(E, E') \otimes \mathbb{Z}_\ell \to \operatorname{Hom}_{\mathbb{F}_q}(T_\ell(E), T_\ell(E'))$$

is an isomorphism, where $\operatorname{Hom}_{\mathbb{F}_q}(E, E')$ denotes the group of isogenies $E \to E'$ defined over $\mathbb{F}_q$ and $\operatorname{Hom}_{\mathbb{F}_q}(T_\ell(E), T_\ell(E'))$ denotes the group of $\mathbb{Z}_\ell$-linear maps that commute with the action of the q-power Frobenius Galois automorphism. In the first paragraph, we showed that this Frobenius action is *scalar* so commuting is automatic, and

$$\operatorname{Hom}_{\mathbb{F}_q}(T_\ell(E), T_\ell(E')) = \operatorname{Hom}(\mathbb{Z}_\ell^2, \mathbb{Z}_\ell^2) \simeq \mathrm{M}_2(\mathbb{Z}_\ell);$$

and $\operatorname{Hom}_{\mathbb{F}_q}(E, E') = \operatorname{Hom}(E, E')$ has $\operatorname{rk}_{\mathbb{Z}} \operatorname{Hom}(E, E') = 4$.

Finally, we can precompose by endomorphisms of O so $\operatorname{Hom}(E, E')$ is a torsion-free $\mathbb{Z}$-module with a right action by O. Let $\psi \in \operatorname{Hom}(E, E')$ be nonzero and let $\psi^\vee : E' \to E$ be the dual isogeny. Then $I := \psi^\vee \operatorname{Hom}(E, E') \subseteq O$ is an integral right O-ideal; since O is a maximal order by Theorem 42.1.9, the right O-ideal I is necessarily invertible (see 23.1.1), and the same then holds for $\operatorname{Hom}(E, E')$ as a right O-module. The same is true as a left $\operatorname{End}(E')$-module, and these two actions commute. □

42.2 Supersingular isogenies

We now investigate the quaternionic endomorphism rings of supersingular elliptic curves in more detail; we use Waterhouse [Wate69, §3] as our main reference. Let E be a supersingular elliptic curve over $F := \mathbb{F}_p^{\mathrm{al}}$, let $\mathcal{O} := \operatorname{End}(E)$, and let $B := \mathcal{O} \otimes \mathbb{Q}$. By Theorem 42.1.9, we have $\operatorname{Ram}(B) = \{p, \infty\}$, and $\mathcal{O} \subseteq B$ is a maximal order. Thus $\operatorname{disc} B = p = \operatorname{discrd} \mathcal{O}$ (recalling section 15.1).

We temporarily and briefly need the language of group schemes, at the level of Waterhouse [Wate69]. The reader who is unfamiliar with this language is advised to skip to 42.2.4 and restrict consideration to left ideals I with $\operatorname{nrd}(I)$ coprime to p (or equivalently, separable isogenies).

42.2.1. Let $I \subseteq \mathcal{O}$ be a nonzero integral left $\mathcal{O}$-ideal. Since $\mathcal{O}$ is maximal, necessarily I is locally principal (in particular, invertible) by Proposition 16.1.2.

We define $E[I] \subseteq E$ to be the scheme-theoretic intersection

$$E[I] := \bigcap_{\alpha \in I} E[\alpha] \tag{42.2.2}$$

where $E[\alpha] = \ker \alpha$ as a group scheme over F.

Accordingly, there exists an isogeny $\phi_I : E \to E_I$ where $E_I = E/E[I]$.

42.2.3. We will not need much about the theory of group schemes except that we can measure the degree of an isogeny via the rank of its kernel. Let $H \leq E(F)$ be a finite F-subgroup scheme, for example, $H := \ker \phi$ for $\phi \colon E \to E'$ an isogeny. Then $H = \operatorname{Spec} A_H$ is affine and A_H is a finite F-algebra; we define the **rank** of H by $\operatorname{rk} H := \dim_F A_H$. In all cases, we have $\operatorname{rk} \ker \phi = \deg \phi$, even when ϕ is inseparable.

This general, scheme-theoretic construction can usually be given plainly, as follows.

42.2.4. If there is a nonzero $\alpha \in I$ giving a *separable* isogeny $\alpha \colon E \to E$, then

$$E[I](F) = \{P \in E(F) : \alpha(P) = 0 \text{ for all } \alpha \in I\}. \tag{42.2.5}$$

We then have the more familiar separable isogeny $\phi_I : E \to E/E[I]$ with $\ker(\phi_I) = E[I]$ [Sil2009, Proposition III.4.12], and $\operatorname{rk} \ker \phi_I = \# \ker \phi_I(F) = \deg \phi_I$.

What remains are inseparable isogenies. Since $\operatorname{Ram} B = \{p, \infty\}$ and $\mathcal{O}$ is maximal, by Theorem 18.1.3 (more generally, see 23.3.19), there is a unique two-sided $\mathcal{O}$-ideal $P \subseteq \mathcal{O}$ of reduced norm p. Then the map $E \to E_P \simeq E^{(p)}$ is the p-Frobenius map, and $\operatorname{rk} \ker \phi_P = p = \deg \phi$ even though $(\ker \phi)(F) = \{0\}$. The equality $P^2 = p\mathcal{O}$ corresponds to the fact that $E[P^2] = E[p]$, and this lies behind the fact that $j(E) \in \mathbb{F}_{p^2}$ as in Proposition 42.1.7.

Accordingly, a left $\mathcal{O}$-ideal I can be written uniquely as $I = P^r I'$ with $\operatorname{nrd}(I')$ coprime to p, and this corresponds to a factorization

$$\phi_I : E \to E_{P^r} \to E_I \tag{42.2.6}$$

with $E_I \simeq E_{P^r}/E_{P^r}[I']$, the isogeny $E \to E_{P^r}$ purely inseparable and the isogeny $E_{P^r} \to E_I$ separable. (This corresponds to the factorization of the extension of function fields into first a separable extension, then a purely inseparable extension [Sil2009, Corollary II.2.12].)

Lemma 42.2.7. *The pullback map*

$$\begin{aligned} \phi_I^*\colon \operatorname{Hom}(E_I, E) &\to I \\ \psi &\mapsto \psi\phi_I \end{aligned} \tag{42.2.8}$$

is an isomorphism of left O-modules.

Proof. The image of $\operatorname{Hom}(E_I, E)$ under precomposition by ϕ_I lands in $\operatorname{End}(E) = O$ and factors through ϕ_I so lands in I by definition. The map ϕ_I^* is an injective homomorphism of abelian groups. It compatible with the left O-action, given by postcomposition on $\operatorname{Hom}(E_I, E)$ and left multiplication on I.

To conclude, we show it is also surjective. Let $\alpha \in I$; then $\alpha(E[I]) = \{0\}$ by construction. If α is separable, then α factors through $\phi_I : E \to E_I$ [Sil2009, Corollary III.4.11]. In general, we factor ϕ_I as in (42.2.6), and then combine the separable case in the previous sentence with the p^r-Frobenius map. □

Finally, we can identify the right module structure as follows.

Lemma 42.2.9. *The ring homomorphism*

$$\begin{aligned} \iota\colon \operatorname{End}(E_I) &\hookrightarrow B \\ \iota(\beta) = \phi_I^{-1}\beta\phi_I &= \frac{1}{\deg\phi_I}(\phi_I^\vee\beta\phi_I) \end{aligned} \tag{42.2.10}$$

is injective and $\iota(\operatorname{End}(E_I)) = O_{\mathsf{R}}(I)$.

Proof. The equality in (42.2.10) is justified in (42.1.4). The content of the lemma follows from the identification in the previous Lemma 42.2.7, by transporting structure: for $\beta \in \operatorname{End}(E_I)$ acting by precomposition, we fill in the diagram

$$\begin{array}{ccc} \operatorname{Hom}(E_I, E) & \xrightarrow{\phi_I^*} & I \\ \Big\downarrow{\scriptstyle\beta} & & \Big\downarrow{\scriptstyle\beta^*} \\ \operatorname{Hom}(E_I, E) & \xrightarrow{\phi_I^*} & I \end{array} \tag{42.2.11}$$

to find that

$$\beta^*(\psi\phi_I) = \psi\beta\phi_I = \psi\phi_I(\phi_I^{-1}\beta\phi_I) \tag{42.2.12}$$

and so ι defines the induced action on I by right multiplication, giving an inclusion $\iota(\operatorname{End}(E_I)) \subseteq O_{\mathsf{R}}(I)$. But $\operatorname{End}(E_I)$ is a maximal order and $O_{\mathsf{R}}(I)$ is an order, so equality holds. □

Next, we show that the isomorphism class of E_I depends only on the left ideal class of I.

Lemma 42.2.13. *If* $J = I\beta \subseteq O$ *with* $\beta \in B^\times$, *then* $E_I \simeq E_J$.

Proof. First, suppose $\beta \in O$. Then

$$E[I\beta] = \{P \in E(F) : \alpha\beta(P) = 0 \text{ for all } \alpha \in I\}.$$

We claim that $\beta E[I\beta] = E[I]$. The containment ($\subseteq$) is immediate. For the containment ($\supseteq$), let $Q \in E[I]$. Since β is surjective (it is nonconstant), there exists $P \in E(F)$ such that $\beta(P) = Q$. Thus for all $\alpha \in I$, we have $(\alpha\beta)(P) = \alpha(Q) = 0$ so $P \in E[I\beta]$. By the claim, we conclude that $\phi_{I\beta} = \phi_I \beta$ and $E_{I\beta} \simeq E_I$.

In general, there exists nonzero $m \in \mathbb{Z}$ such that $m\beta \in O$. By the previous paragraph, we have isomorphisms $E_I \simeq E_{I(m\beta)} = E_{(I\beta)m} \simeq E_{I\beta}$. □

So far, we have shown how to pass from classes of left O-ideals to (isogenous) supersingular elliptic curves via kernels. We can also go in the other direction.

42.2.14. Given a finite subgroup scheme $H \leq E(F)$, we define

$$I(H) := \{\alpha \in O : \alpha(P) = 0 \text{ for all } P \in H\} \subseteq O;$$

then $I(H)$ is a left O-ideal, nonzero because $\#H \in I(H)$.

If $H_1 \leq H_2 \leq E(F)$ are two such subgroups, then $I(H_1) \supseteq I(H_2)$.

Lemma 42.2.15. *If $H_1 \subseteq H_2$ and $I(H_1) = I(H_2)$, then $H_1 = H_2$.*

Proof. Let $\phi_1 : E \to E/H_1$. Factoring, without loss of generality we may assume that ϕ_1 is either separable or purely inseparable. Suppose first that ϕ_1 is separable, and let $n = \#H_2(F)$. By the proof of Theorem 42.1.9, the structure map $O/nO \to \operatorname{End} E[n]$ is faithful. So if $H_2 > H_1$, then there exists $\alpha \in O$ such that $\alpha(H_1) = \{0\}$ but $\alpha(H_2) \neq \{0\}$, so $I(H_2) \neq I(H_1)$. Second, suppose that ϕ_1 is purely inseparable: then $H_1 = \ker \phi_p^{r_1}$ is the kernel of the r_1-power Frobenius for some $r_1 > 0$, and $I(H_1) = P^{r_1}$ as in 42.2.4. Then $p^{r_1} \in I(H_1) = I(H_2)$, so $E \to E/H_2$ is also purely inseparable, and $H_2 = \ker \phi_p^{r_2}$ and $I(H_2) = P^{r_2}$. We conclude $r_1 = r_2$, and then $H_1 = H_2$. □

Proposition 42.2.16. *The following statements hold.*

(a) $\deg \phi_I = \operatorname{nrd}(I)$.
(b) $I(E[I]) = I$.

Proposition 42.2.16 justifies the use of overloaded notation. Our proof follows Waterhouse [Wate69, Theorem 3.15].

Proof. We begin with (a). We first prove it in an illustrative special case. Suppose $I = O\beta$ is a principal left O-ideal. Then $E[I] = E[\beta]$ where $\phi_I = \beta \colon E \to E$, and

$$\deg \beta = \beta\overline{\beta} = \operatorname{nrd}(\beta) = \operatorname{nrd}(I)$$

is the constant term of the (reduced) characteristic polynomial of β.

We now return to the general case. We first show that $\deg(I) \mid \deg \phi_I$. Let $O' := O_{\mathsf{R}}(I) = O_{\mathsf{L}}(I^{-1})$. By Exercise 17.5, there exists $\alpha \in B^\times$ such that $I' = I^{-1}\alpha \subseteq O'$ is in the same right O'-ideal class as I^{-1} and with $\operatorname{nrd}(I')$ coprime to $\deg \phi_I$. Thus

$II' = O\alpha \subseteq O$ is a compatible product and $\alpha \in O$. By the previous paragraph, we have

$$\deg \phi_{II'} = \operatorname{nrd}(II') = \operatorname{nrd}(I)\operatorname{nrd}(I').$$

The map $\phi_{II'}$ factors through ϕ_I, so $\deg \phi_I \mid \deg \phi_{II'}$. Since $\operatorname{nrd}(I')$ is coprime to $\deg \phi_I$ we have $\deg \phi_I \mid \operatorname{nrd}(I)$.

Repeating this argument, we have $\deg \phi_{I'} \mid \operatorname{nrd}(I')$ as well. We combine these to conclude (a). We have

$$(\deg \phi_I)(\deg \phi_{I'}) \mid \operatorname{nrd}(I)\operatorname{nrd}(I') = \deg \phi_{II'}. \tag{42.2.17}$$

But

$$\deg \phi_{II'} = \operatorname{rk} E[II'] \mid (\operatorname{rk} E[I])(\operatorname{rk} E[I']) = (\deg \phi_I)(\deg \phi_{I'}). \tag{42.2.18}$$

Putting together (42.2.17)–(42.2.18), we see that equality holds, so $\deg \phi_I = \operatorname{nrd}(I)$.

Now we prove (b). Let $J = I(E[I])$. Then $I \subseteq J$ since $IE[I] = \{0\}$; thus $E[I] \supseteq E[J]$. At the same time, we have

$$E[J] = \bigcap_{\alpha \in J} E[\alpha] = \bigcap_{\substack{\alpha \in O \\ \alpha E[I] = \{0\}}} E[\alpha] \supseteq \bigcap_{\alpha \in I} E[\alpha] = E[I] \tag{42.2.19}$$

so equality holds and $E[I] = E[J]$. Thus $\deg \phi_I = \deg \phi_J$. By (a), we have

$$\operatorname{nrd}(I) = \deg \phi_I = \deg \phi_J = \operatorname{nrd}(J). \tag{42.2.20}$$

Since $I \subseteq J$, we have $O \subseteq JI^{-1}$. From Proposition 16.4.3 and (42.2.20), we have $\operatorname{nrd}(JI^{-1}) = [O : JI^{-1}] = 1$ so $O = JI^{-1}$ and therefore $I = J$. ∎

Corollary 42.2.21. *For every isogeny $\phi\colon E \to E'$, there exists a left O-ideal I and an isomorphism $\rho\colon E_I \to E'$ such that $\phi = \rho\phi_I$. Moreover, for every maximal order $O' \subseteq B$, there exists E' such that $O' \simeq \operatorname{End}(E')$.*

Proof. Let H be the scheme-theoretic kernel of ϕ. Then $H \subseteq E[I(H)]$, so $\phi_{I(H)}$ factors through ϕ with $\phi_{I(H)} = \rho\phi$ for some isogeny $\rho\colon E_{I(H)} \to E'$. But $I(H) = I(E[I(H)])$ by Proposition 42.2.16, so $H = E[I(H)]$ by Lemma 42.2.15. Thus $\deg \phi_{I(H)} = \deg \phi$, and so $\deg \rho = 1$ and ρ is an isomorphism, with $\phi = \rho^{-1}\phi_{I(H)}$. The second statement follows similarly using a connecting ideal between orders (see section 17.4). ∎

We may now compare endomorphisms analogously to Lemma 42.2.7.

Lemma 42.2.22. *Let $I, I' \subseteq O$ be nonzero integral left O-ideals. Then the natural map*

$$\operatorname{Hom}(E_I, E)\operatorname{Hom}(E_{I'}, E_I) \to \operatorname{Hom}(E_{I'}, E)$$

is bijective, giving a further bijection

$$\begin{aligned} \operatorname{Hom}(E_{I'}, E_I) &\to (I' : I)_{\mathsf{R}} = I^{-1}I' \\ \psi &\mapsto \phi_I^{-1}\psi\phi_{I'}. \end{aligned} \tag{42.2.23}$$

Proof. By Lemma 42.2.7, we have $\operatorname{Hom}(E_I, E)\phi_I = I$. By Proposition 42.2.16, we have $m = \deg \phi_I = \operatorname{nrd}(I)$. The left ideal $I \subseteq O$ is invertible thus $m = \operatorname{nrd}(I) \in I\overline{I}$, hence there exist finitely many $\alpha_i, \beta_i \in \operatorname{Hom}(E_I, E)$ such that

$$[m] = \sum_i (\alpha_i\phi_I)(\beta_i\phi_I)^\vee = \sum_i (\alpha_i\phi_I)(\phi_I^\vee\beta_i^\vee) = \sum_i \alpha_i[m]\beta_i^\vee \tag{42.2.24}$$

therefore

$$[1] = \sum_i \alpha_i\beta_i^\vee \in \operatorname{End}(E). \tag{42.2.25}$$

For $\psi \in \operatorname{Hom}(E_{I'}, E)$, postcomposing with (42.2.25) gives

$$\psi = \sum_i \alpha_i(\beta_i^\vee\psi) \in \operatorname{Hom}(E_I, E)\operatorname{Hom}(E_{I'}, E_I) \tag{42.2.26}$$

so the natural injective map is bijective. This gives

$$I\phi_I^{-1}\operatorname{Hom}(E_{I'}E_I)\phi_{I'} = I' \tag{42.2.27}$$

and thereby the bijective map (42.2.23), using Exercise 16.11 for the relationship to the colon ideal.. □

42.3 Equivalence of categories

We now show that the association from supersingular elliptic curves to right ideals is an equivalence of categories. We recall that F is an algebraically closed field with $p = \operatorname{char} F$ prime.

42.3.1. Let E_0 be a supersingular elliptic curve over $F := \mathbb{F}_p^{\text{al}}$; it will serve the role as a *base object*. Let $O_0 := \operatorname{End}(E_0)$ and $B_0 := O_0 \otimes \mathbb{Q}$.

Theorem 42.3.2. *The association* $E \mapsto \operatorname{Hom}(E, E_0)$ *is functorial and defines and equivalence between the category of*

supersingular elliptic curves over F, under isogenies

and

invertible left O_0-modules, under nonzero left O_0-module homomorphisms.

Remark 42.3.3. Written this way, the functor $\operatorname{Hom}(-, E_0)$ is contravariant. One can equally well take $\operatorname{Hom}(E_0, -)$ to get a covariant functor with right O_0-modules; using the standard involution, these are seen to contain the same content.

Proof. To begin, we need to show $\operatorname{Hom}(-, E_0)$ is a functor. The association $E \mapsto \operatorname{Hom}(E, E_0)$ makes sense on objects by Lemma 42.1.11. On morphisms, to an isogeny $\phi\colon E \to E'$ we associate

$$\begin{aligned} \phi^*\colon \operatorname{Hom}(E', E_0) &\to \operatorname{Hom}(E, E_0) \\ \psi &\mapsto \psi\phi. \end{aligned} \tag{42.3.4}$$

The map ϕ^* is a homomorphism of left O_0-modules, since it is compatible with postcomposition with $O_0 = \operatorname{End}(E)$, so $\operatorname{Hom}(-, E_0)$ is functorial.

Next, we claim that $\operatorname{Hom}(-, E_0)$ is essentially surjective. Let I be an invertible left O_0-module. Tensoring with $\mathbb{Q}$ we get an injection $I_0 \hookrightarrow I_0 \otimes \mathbb{Q} \simeq B_0$, so up to isomorphism of left O_0-modules, we may suppose $I \subseteq B_0$. Scaling by an integer, we may suppose $I \subseteq O_0$ is a left O_0-ideal. Let $E_I = E/E[I]$. By Lemma 42.2.7, we have $\operatorname{Hom}(E_I, E_0) \simeq I$ as left O_0-modules, as desired.

Finally, we show that $\operatorname{Hom}(-, E_0)$ is fully faithful, i.e., the map

$$\begin{aligned} \operatorname{Hom}(E, E') &\to \operatorname{Hom}(\operatorname{Hom}(E', E_0), \operatorname{Hom}(E, E_0)) \\ \phi &\mapsto \phi^* \end{aligned}$$

(the former as isogenies, the latter as homomorphisms of left O_0-modules) is bijective. This bijectivity is made plain by an application of Corollary 42.2.21: there is a left O_0-ideal I such that $E \simeq E_{0,I}$; applying this isomorphism, we may suppose without loss of generality that $E = E_{0,I}$. Then by Lemma 42.2.7, we have $I = \operatorname{Hom}(E_{0,I}, E_0)\phi_{0,I}$. Repeat with E' and I'. Then after these identifications, we are reduced to the setting of Lemma 42.2.22 (with the location of the prime swapped): the map

$$\begin{aligned} \operatorname{Hom}(E_{0,I}, E_{0,I'}) &\to (I' : I)_{\mathsf{R}} = I'^{-1} I \\ \psi &\mapsto \phi_{0,I'}^{-1} \psi \phi_{0,I}. \end{aligned} \tag{42.3.5}$$

is indeed bijective. □

Remark 42.3.6. See also Kohel [Koh96, Theorem 45], where the categories are enriched with a Frobenius morphism.

Corollary 42.3.7. (Deuring correspondence). *There is a bijection between isomorphism classes of supersingular elliptic curves over F and the left class set* $\operatorname{Cls}_{\mathsf{L}} O_0$. *Under this bijection, if $E \leftrightarrow [I]$, then* $\operatorname{End}(E) \simeq O_{\mathsf{R}}(I)$ *and* $\operatorname{Aut}(E) \simeq O_{\mathsf{R}}(I)^\times$.

Proof. Take isomorphism classes on both sides of the equivalence in Theorem 42.3.2, and compare endomorphism groups and automorphism groups. (We had to work with left O_0-modules in the equivalence of categories, but each isomorphism class of objects is represented by a left O_0-ideal $I \subseteq B$.) □

42.3.8. From the Eichler mass formula and Corollary 42.3.7 (swapping left for right, as in Remark 42.3.3), we conclude that

$$\sum_{[E]} \frac{1}{\#\operatorname{Aut}(E)} = \sum_{[I] \in \operatorname{Cls}_{\mathsf{R}} O} \frac{1}{\# O_{\mathsf{L}}(I)^\times} = \frac{p-1}{24} \tag{42.3.9}$$

where the sum on the left is over isomorphism classes of supersingular elliptic curves over $F = \mathbb{F}_p^{\mathrm{al}}$.

Similarly, from the Eichler class number formula (Theorem 30.1.5), the number of isomorphism classes of supersingular elliptic curves over F is equal to

$$\frac{p-1}{12} + \frac{\epsilon_2}{4}\left(1 - \left(\frac{-4}{p}\right)\right) + \frac{\epsilon_3}{3}\left(1 - \left(\frac{-3}{p}\right)\right).$$

Remark 42.3.10. We can generalize this setup slightly as follows. Let $M \in \mathbb{Z}_{>0}$ be coprime to p, and let $C_0 \le E_0(F)$ be a cyclic subgroup of order M. Then $\operatorname{End}(E_0, C_0) \simeq O_0(M)$ is an Eichler order of level M and reduced discriminant pM in B_0. In a similar way as above, one can show that $\operatorname{Hom}(-, (E_0, C_0))$ defines an equivalence of categories between the category of supersingular elliptic curves equipped with a cyclic M-isogeny (under isogenies identifying the cyclic subgroups), to the category of left invertible $O_0(M)$-modules (under homomorphisms). The mass formula now reads

$$\sum_{[(E,C)]} \frac{1}{\#\operatorname{Aut}(E,C)} = \sum_{[I]\in \operatorname{Cls} O_0(M)} \frac{1}{\# O_{\mathsf{L}}(I)} = \frac{p-1}{24}\psi(M).$$

One can also consider instead the category of cyclic M-isogenies $\phi\colon E \to E'$.

Example 42.3.11. Consider $p = 11$. The algebra $B = \left(\dfrac{-1,-11}{\mathbb{Q}}\right)$ has discriminant 11 and the maximal order $O = \mathbb{Z}\langle i, (1+j)/2\rangle$. We have $\#\operatorname{Cls} O = 2$, with the nontrivial class represented by the ideal I generated by 2 and $1 + i(1+j)/2$.

We have $O^\times = \langle i\rangle$ of order 4 and $O_{\mathsf{L}}(I) = \langle 1/2 - i(1+j)/4\rangle$ of order 6, and indeed $1/4 + 1/6 = 10/24 = 5/12$. The two supersingular curves modulo 11 are the ones with j-invariants 0 and $1728 \equiv 1 \pmod{11}$, and $\operatorname{End}(E) \simeq O$ if $j(E) = 1728$ whereas for $\operatorname{End}(E') \simeq O'$ we have $\operatorname{Hom}(E, E') \simeq I$, in other words, $E' \simeq E/E[I]$.

Example 42.3.12. We return to Example 41.1.2. The order $O = O_1$ is the endomorphism ring of the elliptic curve $E_1\colon y^2 = x^3 - x$ with $j(E_1) = 1728 \equiv 3 \pmod{23}$, and similarly $E_2\colon y^2 = x(x-1)(x+2)$ with $j(E_2) = 19$ and $E_3\colon y^2 = x^3 + 1$ with $j(E_3) = 0$. We have $2w_i = \#\operatorname{Aut}(E_i)$ is the order of the automorphism group of E_i. And the p-Brandt graph is the graph of p-isogenies among the three supersingular elliptic curves over $\overline{\mathbb{F}}_{23}$.

42.3.13. Finally, and most importantly, in the above correspondence the entries of the Brandt matrix $T(n)$ have meaning as counting isogenies. For n coprime to p, the entry $T(n)_{ij}$ is equal to the number of subgroups $H \le E_i(F)$ such that $E_i/H \simeq E_j$. This statement is just a translation of Lemma 42.2.22. Gross [Gro87] gives a beautiful and essentially self-contained presentation of the results of the previous chapter in the special case that $\operatorname{disc} B = p$.

Remark 42.3.14. The approach via supersingular elliptic curves connects back in another way: Serre [Ser96] gives an alternative approach to modular forms modulo p in a letter to Tate: one evaluates classical modular forms at supersingular elliptic curves and then relates these to quaternionic modular forms modulo p.

42.4 Supersingular endomorphism rings

In this section, we give a second categorical perspective, giving a base-object free refinement of Corollary 42.3.7 following Ribet [Rib1989, p. 360–361] (who credits Mestre–Oesterlé). To get there, we need to deal with a small subtlety involving the field of definition (fixed by keeping track of extra data). Recall that $F = \mathbb{F}_p^{\mathrm{al}}$.

Lemma 42.4.1. *Let O be a maximal order. Then there exist one or two supersingular elliptic curves E up to isomorphism over F such that* $\mathrm{End}(E) \simeq O$*. There exist two such elliptic curves if and only if $j(E) \in \mathbb{F}_{p^2} \smallsetminus \mathbb{F}_p$ if and only if the unique two-sided ideal of O of reduced norm p is not principal.*

Proof. In Corollary 42.2.21, we proved that there is always at least one supersingular elliptic curve E with $\mathrm{End}(E) \simeq O$ using a connecting ideal. We now elaborate on this point, refining our count.

By Corollary 42.3.7, the isomorphism classes of supersingular elliptic curves are in bijection with the left class set $\mathrm{Cls}_{\mathsf{L}}\, O_0$; their endomorphism rings are then given by $\mathrm{End}(E) \simeq O_{\mathsf{R}}(I)$ for $[I] \in \mathrm{Cls}_{\mathsf{L}}\, O_0$. By Lemma 17.4.13 (interchanging left for right), the map

$$\begin{aligned} \mathrm{Cls}_{\mathsf{L}}\, O_0 &\to \mathrm{Typ}\, O_0 \\ [I]_{\mathsf{L}} &\mapsto \text{class of } O_{\mathsf{R}}(I) \end{aligned} \tag{42.4.2}$$

is a *surjective* map of sets. The connecting ideals are precisely the fibers of this map, and by the bijection of Corollary 42.3.7, there is a bijection between the set of supersingular elliptic curves E with $\mathrm{End}(E) \simeq O$ and the fiber of this map over the isomorphism class of O.

We now count these fibers. We recall Theorem 18.1.3 with $D = p$ and the text that follows (interchanging left for right): the fibers are given by the quotient group $\mathrm{PIdl}\, O \backslash \mathrm{Idl}\, O$ of the two-sided invertible fractional two-sided O-ideals by the subgroup of principal such ideals. There is a surjection $\mathrm{Pic}(O) \to \mathrm{PIdl}\, O \backslash \mathrm{Idl}\, O$ and $\mathrm{Pic}(O) \simeq \mathbb{Z}/2\mathbb{Z}$ is generated by the unique maximal two-sided ideal P of reduced norm P. The class of P in the quotient is trivial if and only if $P = O\pi$ is principal.

To conclude we recall 42.2.4: the Frobenius map is the map $E \to E_P \simeq E^{(p)}$. So P is principal if and only if $E^{(p)} \simeq E$ if and only if $j(E) = j(E^{(p)}) = j(E)^p$ if and only if $j(E) \in \mathbb{F}_p$. ∎

We dig into this issue a bit further.

42.4.3. Let E be a supersingular elliptic curve over $\mathbb{F}_p^{\mathrm{al}}$. Let ω be a nonzero invariant differential on E. Then there is a ring homomorphism

$$\begin{aligned} a\colon\ \mathrm{End}(E) &\to \mathbb{F}_p^{\mathrm{al}} \\ \phi &\mapsto a_\phi, \qquad \text{where } \phi^*\omega = a_\phi \omega \end{aligned} \tag{42.4.4}$$

(see Silverman [Sil2009, Corollary III.5.6]) independent of the choice of ω.

In light of 42.4.3, we make the following definitions. Let B be a quaternion algebra over $\mathbb{Q}$ of discriminant $\mathrm{disc}\, B = p$; such an algebra B is unique up to isomorphism. Let $O \subseteq B$ be a maximal order in B; then $\mathrm{discrd}\, O = p$.

Definition 42.4.5. An **orientation** of O is a ring homomorphism $O \to \mathbb{F}_p^{\mathrm{al}}$.

42.4.6. We claim that there are two possible orientations of O. In fact, $P = [O, O]$ is the commutator ideal (cf. Exercise 13.7) and an orientation factors through the commutator. Since $\mathrm{nrd}(P) = p$, localizing we have $O/P \simeq O_p/P_p \simeq \mathbb{F}_{p^2}$ by Theorem 13.3.11(c). The claim follows as there are two possible inclusions $\mathbb{F}_{p^2} \hookrightarrow \mathbb{F}_p^{\mathrm{al}}$.

The notion of isomorphism of oriented maximal orders is evident.

Definition 42.4.7. An isomorphism of oriented maximal orders from (O,ζ) to (O',ζ') is an isomorphism of orders $\phi : O \to O'$ such that $\zeta'\phi = \zeta$.

We define the set of **reduced isomorphisms** to be $\mathrm{Isom}(E,E')/\{\pm 1\}$.

Proposition 42.4.8. *The association* $E \mapsto (\mathrm{End}(E) \subseteq \mathrm{End}(E)_{\mathbb{Q}}, a)$ *is functorial and induces an equivalence from the category of*

supersingular elliptic curves over $\mathbb{F}_p^{\mathrm{al}}$, under reduced isomorphisms

to the category of

oriented maximal orders $(O \subseteq B, \zeta)$
in a quaternion algebra B of discriminant p,
under isomorphisms.

In the latter category, we do not choose the representative of the isomorphism class of quaternion algebra of discriminant p; it is tagging along only to provide a quaternionic wrapper for the order.

Proof. The association has the right target by Theorem 42.1.9 for the order and 42.4.3 for the orientation. This association is (covariantly) functorial with respect to isomorphisms. Indeed, if $\psi : E \xrightarrow{\sim} E'$ is an isomorphism of elliptic curves, then we have an induced isomorphism

$$\begin{aligned} \mathrm{End}(E) &\to \mathrm{End}(E') \\ \phi &\mapsto \psi\phi\psi^{-1}. \end{aligned} \tag{42.4.9}$$

that is compatible with composition. The isomorphism (42.4.9) is also compatible with orientations, as follows. Let ω' be a nonzero invariant differential on E'; then $\psi^*\omega'$ is so on E. Thus for all $\phi \in \mathrm{End}(E)$, we have

$$\begin{aligned} a_\phi \psi^*\omega' = \phi^*\psi^*\omega' &= \psi^*(\psi^*)^{-1}\phi^*\psi^*\omega' = \psi^*(\psi\phi\psi^{-1})^*\omega' \\ &= \psi^*(a'_{\psi\phi\psi^{-1}}\omega') = a'_{\psi\phi\psi^{-1}}\psi^*\omega' \end{aligned} \tag{42.4.10}$$

so $a'_{\psi\phi\psi^{-1}} = a_\phi$, which is the desired compatibility.

The functor is essentially surjective, which is to say that every oriented maximal order arises up to isomorphism: that every maximal order arises is a consequence of Corollary 42.2.21, and that the orientation may be so chosen corresponds to applying the Frobenius morphism, by 42.2.4 and 42.4.6.

Finally, we show the map is fully faithful, which is to say the map of finite sets

$$\mathrm{Isom}(E,E')/\{\pm 1\} \to \mathrm{Isom}((\mathrm{End}\,E,\zeta),(\mathrm{End}\,E',\zeta'))$$

from (42.4.9) is bijective. Any two reduced isomorphisms on the left differ by an automorphism of E, and the same on the right, so it suffices to show that the map

$$\begin{aligned} \mathrm{Aut}(E)/\{\pm 1\} &\to \mathrm{Aut}((\mathrm{End}(E),\zeta)) \\ \nu &\mapsto (\phi \mapsto \nu\phi\nu^{-1}) \end{aligned} \tag{42.4.11}$$

is bijective. Let $O = \mathrm{End}(E)$, so $\mathrm{Aut}(E) \simeq O^\times$. Then $\mathrm{Aut}(O) \simeq N_{B^\times}(O)/\mathbb{Q}^\times$, and

$$1 \to O^\times/\{\pm 1\} \to \mathrm{Aut}(O) \to \mathrm{AL}(O) \to 1$$

where the Atkin–Lehner group $\mathrm{AL}(O)$ is nontrivial (and isomorphic to $\mathbb{Z}/2\mathbb{Z}$) if and only if $P = \pi O$ is principal; but conjugation by π acts nontrivially on O/P and fails to commute with ζ so does *not* act by an automorphism of $(\mathrm{End}(E), \zeta)$; thus

$$\mathrm{Aut}(E)/\{\pm 1\} \simeq O^\times/\{\pm 1\} \simeq \mathrm{Aut}(O, \zeta).$$

(We did not need to choose a base object in order to define this equivalence!) □

Corollary 42.4.12. *There is a bijection between isomorphism classes of supersingular elliptic curves over F and oriented maximal orders in a quaternion algebra B of discriminant p.*

Proof. Again we take isomorphism classes of objects in Proposition 42.4.8. □

Exercises

1. Let E, E' be elliptic curves over F and suppose E, E' are isogenous (necessarily by a nonzero isogeny). Show that E is supersingular if and only if E' is supersingular. *[Hint: show* $\dim_{\mathbb{Q}} \mathrm{End}(E_{F^{\mathrm{al}}})_{\mathbb{Q}} = \dim_{\mathbb{Q}} \mathrm{End}(E'_{F^{\mathrm{al}}})_{\mathbb{Q}}$; *or show that* $\deg_i([p]) = \deg_i([p]')$ *where* $[p], [p]'$ *are multiplication by p on E, E'.]*

2. Let E be an elliptic curve over F with $\mathrm{char}\, F = p$. Show that for all $\phi, \psi \in \mathrm{End}(E)$, we have

$$\deg_i(\phi\psi) = \deg_i(\phi)\deg_i(\psi)$$
$$\deg_i(\phi + \psi) \geq \min\{\deg_i \phi, \deg_i \psi\}.$$

Conclude that $|\phi| = 1/\deg_i(\phi)$ defines a nonarchimedean absolute value on $\mathrm{End}(E)_{(p)}$.

3. In this exercise, we give an alternate "hands on" proof of Lemma 42.2.9.

Let E be a supersingular elliptic curve over $F = \mathbb{F}_p^{\mathrm{al}}$, let $O = \mathrm{End}(E)$ and let $B = O \otimes \mathbb{Q}$. Let $I \subseteq O$ be a nonzero integral left O-ideal, and let $\phi_I : E \to E_I$ where $E_I = E/E[I]$. Consider the pullback isomorphism $\phi_I^* \colon \mathrm{Hom}(E_I, E) \to I$ by $\psi \mapsto \psi\phi_I$ in (42.2.8).

(a) Show that ϕ^* induces an isomorphism of O-module endomorphism rings

$$\rho\colon \mathrm{End}(\mathrm{Hom}(E_I, E)) \xrightarrow{\sim} \mathrm{End}(I)$$
$$\alpha \mapsto \phi^*\alpha\phi^{*-1}.$$

(b) Show that $\mathrm{End}(I) = O_{\mathsf{R}}(I)^{\mathrm{op}}$.
(c) Show that $\mathrm{End}(\mathrm{Hom}(E_I, E)) = \mathrm{End}(E_I)^{\mathrm{op}}$. *[Hint:* $\mathrm{End}(E_I)$ *is a maximal order, so the natural inclusion is an equality.]*
(d) Conclude that ρ^{op} induces an isomorphism $\mathrm{End}(E_I) \xrightarrow{\sim} O_{\mathsf{R}}(I)$.

(e) Show that ρ^{op} is the map ι in (42.2.10).

4. Let B be a quaternion algebra over $\mathbb{Q}$ with disc $B = p$ prime. Recall the definition of the Brandt matrix (e.g. (41.1.1), and more generally 41.2.2).

   Let I_i for $i = 1, \dots, h$ be the representatives of the left class set $\operatorname{Cls}_{\mathsf{L}} \mathcal{O}$, with $h = \# \operatorname{Cls}_{\mathsf{L}} \mathcal{O}$. Let E_i be the supersingular elliptic curves over $\mathbb{F}_p^{\text{al}}$ corresponding to I_i in Corollary 42.3.7.

   (a) For every $m \geq 1$, show that $T(m)_{ij}$ is equal to the number of subgroup schemes C of order m in E_j such that $E_j/C \simeq E_i$.
   (b) Show that $T(p)$ is a permutation matrix of order dividing 2 and that $T(p^r) = T(p)^r$ for all $r \geq 1$.
   (c) Show that E_i is conjugate by an element of $\operatorname{Aut}(\mathbb{F}_p^{\text{al}})$ to E_j if and only if $i = j$ or $T(p)_{ij} = 1$. Conclude that the number of elliptic curves E_i defined over $\mathbb{F}_p$ is equal to $\operatorname{tr} T(p)$.

This chapter is licensed under the terms of the Creative Commons Attribution-NonCommercial 4.0 International License (http://creativecommons.org/licenses/by-nc/4.0/), which permits any noncommercial use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license and indicate if changes were made.

The images or other third party material in this chapter are included in the chapter's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the chapter's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.

Chapter 43

QM abelian surfaces

In this final chapter, we consider quotients of the upper half-plane by quaternionic unit groups as generalizations of such quotients from the matrix group (Chapter 40), realizing them as moduli spaces for abelian surfaces with quaternionic multiplication. This chapter can be seen as a culminating application of all of the parts of this book, and for that reason, is necessarily more advanced. Concepts are reviewed in the attempt to be self-contained, but additional background in algebraic and arithmetic geometry is suggested.

43.1 ▹ QM abelian surfaces

Recall (40.1.1) that the curve $\mathrm{SL}_2(\mathbb{Z})\backslash\mathbf{H}^2$ parametrizes complex elliptic curves up to isomorphism: to $\tau \in \mathbf{H}^2$, we associate the lattice $\Lambda_\tau := \mathbb{Z} + \mathbb{Z}\tau$ and the elliptic curve $E_\tau := \mathbb{C}/\Lambda_\tau$, and the association

$$\begin{aligned} \mathrm{SL}_2(\mathbb{Z})\backslash\mathbf{H}^2 &\leftrightarrow \{\text{Complex elliptic curves up to isomorphism}\} \\ \mathrm{SL}_2(\mathbb{Z})\tau &\mapsto [E_\tau] \end{aligned} \tag{43.1.1}$$

is bijective. Moreover, we have a biholomorphic map $j : \mathrm{SL}_2(\mathbb{Z})\backslash\mathbf{H}^2 \to \mathbb{C}$, which is to say, two complex elliptic curves are isomorphic if and only if they have the same j-invariant. We compactify to $X := \mathrm{SL}_2(\mathbb{Z})\backslash\mathbf{H}^{2*}$ by including the cusp at ∞.

As in section 38.1, we are led to seek a generalization of (43.1.1), replacing $B = \mathrm{M}_2(\mathbb{Q})$ with a quaternion algebra. To this end, let B be an indefinite quaternion algebra over $\mathbb{Q}$ of discriminant D, let $O \subset B$ be a maximal order, and let

$$\iota_\infty : B \to B \otimes_{\mathbb{Q}} \mathbb{R} \simeq \mathrm{M}_2(\mathbb{R})$$

be an embedding (explicitly, we may take (38.1.1)). The order O is unique up to conjugation in B (by strong approximation) and similarly the embedding ι_∞ is unique up to conjugation in $\mathrm{M}_2(\mathbb{R})$, so these choices are harmless. Let

$$\Gamma^1(O) := \iota_\infty(O^1)/\{\pm 1\} \le \mathrm{PSL}_2(\mathbb{R}).$$

© The Author(s) 2021

J. Voight, *Quaternion Algebras*, Graduate Texts in Mathematics 288,
https://doi.org/10.1007/978-3-030-56694-4_43

The quotient $\Gamma^1(O)\backslash\mathbf{H}^2$ is compact when $B \not\simeq \mathrm{M}_2(\mathbb{Q})$; for uniformity, we define

$$X^1 := \Gamma^1(O)\backslash\mathbf{H}^{2(*)},$$

where $\mathbf{H}^{2(*)} = \mathbf{H}^{2*}, \mathbf{H}^2$ according as $D = 1$ or $D > 1$. Then X^1 is a good (compact) complex 1-orbifold.

We may then ask: what does X^1 parametrize? The answer is, roughly: X^1 parametrizes complex abelian surfaces with endomorphisms by O. The correspondence itself is as pleasingly simple as for elliptic curves (43.1.1). To a point $\tau \in \mathbf{H}^2$, we associate

$$\begin{aligned} \Lambda_\tau &:= \iota_\infty(O)\begin{pmatrix}\tau\\1\end{pmatrix} \subseteq \mathbb{C}^2 \\ A_\tau &:= \mathbb{C}^2/\Lambda_\tau \\ \iota_\tau &:= \iota_\infty \colon O \hookrightarrow \mathrm{End}(A_\tau) \end{aligned} \tag{43.1.2}$$

Then A_τ is a complex torus of dimension 2 and ι_τ is an injective ring homomorphism, realizing endomorphisms of A_τ by O.

However, there are a number of technical points required to make this completely precise. We quickly survey the theory of complex abelian varieties in section 43.4. One basic fact of life is that not every complex torus has enough meromorphic functions to give it the structure of a complex abelian variety embedded in projective space. One needs a *polarization* given by a *Riemann form*, and the simplest polarizations are the *principal* polarizations. (One can think of this rigidification as the difference between a genus 1 curve and an elliptic curve, where the genus 1 curve is equipped with a point.) A principal polarization defines positive involution on the endomorphism ring, called the *Rosati involution*.

This rigidification is matched on the quaternion order: a **principal polarization** on O is an element $\mu \in O$ such that $\mu^2 + D = 0$. Every (maximal) order has a principal polarization, and the involution $\alpha \mapsto \alpha^* = \mu^{-1}\overline{\alpha}\mu$ is a positive involution on O. A **quaternionic multiplication (QM) structure** by (O, μ) on a principally polarized complex abelian surface A is an injective ring homomorphism $O \hookrightarrow \mathrm{End}(A)$ that respects the positive involutions on B and $\mathrm{End}(A)_{\mathbb{Q}}$.

The happy fact is that A_τ as defined in (43.1.2) has via μ a principal polarization and thereby QM by (O, μ). In other words, the choice of the QM structure determines a canonical principal polarization: but it gives a finite amount of additional information, as there will in general be more than one QM structure on a principally polarized abelian surface. In many cases, these structures can be understood in terms of the Atkin–Lehner group

$$\mathrm{AL}(O) = N_{B^\times}(O)/\mathbb{Q}^\times O^\times \simeq \prod_{p \mid D} \mathbb{Z}/2\mathbb{Z} \tag{43.1.3}$$

acting by automorphisms of X^1.

In any event, the main result of this chapter (Main Theorem 43.6.14) is that this association is bijective.

Main Theorem 43.1.4. *The map*

$$\begin{aligned} \Gamma^1(O)\backslash \mathbf{H}^2 &\leftrightarrow \left\{ \begin{array}{c} (A,\iota) \text{ principally polarized} \\ \text{complex abelian surfaces} \\ \text{with QM by } (O,\mu) \\ \text{up to isomorphism} \end{array} \right\} \\ \Gamma^1(O)\tau &\mapsto [(A_\tau, \iota_\tau)] \end{aligned} \tag{43.1.5}$$

is a bijection.

This main theorem generalizes (43.1.1): indeed, we may take $B = \mathrm{M}_2(\mathbb{Q}) \supset O = \mathrm{M}_2(\mathbb{Z})$ and $\mu = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$, and we find that $A_\tau \simeq E_\tau^2$ as principally polarized abelian surfaces.

One feature that makes this theory even more appealing is that abelian surfaces arise naturally as Jacobians of genus 2 curves via the Abel–Jacobi map: this motivates much of the theory, so we begin with it in section 43.3. In particular, there are functions called *Igusa invariants* analogous to the elliptic j-function that record the isomorphism class of a principally polarized abelian surface.

43.1.6. We then define modular forms as for the classical modular group. Let $k \in 2\mathbb{Z}_{\geq 0}$. A map $f : \mathbf{H}^2 \to \mathbb{C}$ is **weight k-invariant** under $\Gamma = \Gamma^1(O)$ if

$$f(\gamma z) = (cz+d)^k f(z) \quad \text{for all } \gamma = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in \Gamma. \tag{43.1.7}$$

A **modular form for Γ of weight k** is a holomorphic function $f : \mathbf{H}^2 \to \mathbb{C}$ that is weight k invariant and is holomorphic at ∞, if $\Gamma = \mathrm{PSL}_2(\mathbb{Z})$. Let $M_k(\Gamma)$ be the $\mathbb{C}$-vector space of modular forms for Γ. Then $M_k(\Gamma)$ is a finite-dimensional $\mathbb{C}$-vector space, and by a similar contour integration as in the proof of Proposition 40.3.4, $\dim_{\mathbb{C}} M_k(\Gamma)$ can be expressed in terms of k and the signature of Γ. And

$$M(\Gamma) := \bigoplus_{k \in 2\mathbb{Z}_{\geq 0}} M_k(\Gamma) \tag{43.1.8}$$

has the structure of a graded $\mathbb{C}$-algebra under multiplication. (When $D > 1$, there are no cusps, so vacuously all modular forms are cusp forms.)

It would not be unreasonable for us to have started the book here, with this topic at front and center. In this chapter, we will do our best to treat the complex analytic theory in as complete and self-contained a manner as possible, but this is really just the beginning of the subject, one that is rich, deep, and complicated—worthy of a book all to itself. For example, the following result is fundamental.

Theorem 43.1.9 (Shimura [Shi67, p. 58]). *There exists a projective nonsingular curve X^1 defined over $\mathbb{Q}$ and a biholomorphic map*

$$\varphi : \Gamma^1(O)\backslash \mathbf{H}^2 \xrightarrow{\sim} X^1(\mathbb{C}).$$

The curve X^1 over $\mathbb{Q}$ coarsely represents the functor from schemes over $\mathbb{Q}$ to sets whose values are isomorphism classes of QM abelian schemes, suitably defined. Moreover, the map φ respects the field of definition and Galois action on certain special points called **CM points** on $\Gamma^1(O)\backslash\mathbf{H}^2$ obtained as fixed points of elements $\nu \in B^\times$ with $\mathbb{Q}(\nu)$ an imaginary quadratic field. As a result, the curve X^1 is canonical, uniquely characterized up to isomorphism, and is so called the **canonical model**. We give some indications of this result by example in the next section and more generally in section 43.8.

43.2 ▹ QM by discriminant 6

For concreteness, before embarking on our general treatment, we consider in this section an illustrative example and one of special interest; it is well-studied and beloved by quaternionic practitioners, see Remark 43.2.21 for further reference.

Let $B = \left(\dfrac{-1, 3}{\mathbb{Q}}\right)$ be the quaternion algebra of discriminant 6 studied in sections 37.8–37.9. As in 37.8.12, we have a maximal order

$$O = \mathbb{Z} \oplus \mathbb{Z}i \oplus \mathbb{Z}j \oplus \mathbb{Z}k, \quad k = \frac{1+i+j+ij}{2}$$

with $k^2 - k - 1 = 0$, and an embedding

$$\begin{aligned} \iota_\infty \colon B &\hookrightarrow \mathrm{M}_2(\mathbb{R}) \\ i, j &\mapsto \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}, \begin{pmatrix} \sqrt{3} & 0 \\ 0 & -\sqrt{3} \end{pmatrix} \end{aligned}$$

Let $\Gamma^1 = \iota_\infty(O^1)/\{\pm 1\} \le \mathrm{PSL}_2(\mathbb{R})$ and $X^1 = \Gamma^1\backslash\mathbf{H}^2$. We computed a compact Dirichlet fundamental domain ⌑ for Γ^1 in 37.9.4, with $\mu(⌑) = 2\pi/3$. Further, we saw explicitly in 37.9.10 (and again by formula in Example 39.4.21) that Γ^1 has signature $(0; 2, 2, 3, 3)$; that is, X^1 has topological genus $g = 0$ and there are 4 cone points, two points $z_2, z_2' \in$ ⌑ with stabilizer of order 2 and two $z_3, z_3' \in$ ⌑ with order 3 stabilizer.

As in Chapter 40, to exhibit a model for X^1 we seek modular forms, indeed, we now describe the full graded ring of (even weight) modular forms (43.1.8). We will use the following essential proposition.

Proposition 43.2.1. *The following statements hold.*

(a) *Let $f \colon \mathbf{H}^2 \to \mathbb{C}$ be a nonzero meromorphic modular form of weight k for Γ^1, not identically zero. Then*

$$\sum_{\Gamma^1 z \in \Gamma^1\backslash\mathbf{H}^2} \frac{1}{\#\,\mathrm{Stab}_{\Gamma^1}(z)} \,\mathrm{ord}_z(f) = \frac{k}{6}.$$

(b) *We have*

$$\dim_{\mathbb{C}} M_k(\Gamma^1) = \begin{cases} 1, & \text{if } k = 0; \\ 0, & \text{if } k = 2; \\ 1 - k + 2\lfloor k/4 \rfloor + 2\lfloor k/3 \rfloor, & \text{if } k \ge 4. \end{cases} \qquad (43.2.2)$$

Proof. See Theorem 43.9.4; for the purposes of this introduction, we provide a sketch to tide the reader over. For (a), we argue just as in Proposition 40.3.4: we integrate $\mathrm{dlog}\, f = \mathrm{d}f/f$ over the boundary of the fundamental domain ⌑ and use the identification of sides provided by rotation at their fixed points (elliptic vertices), reversing the direction of the path so the contributions cancel, and we are left again to sum angles. The details are requested in Exercise 43.7. For (b), we can get upper bounds on the dimension using (a), but to provide lower bounds we need to exhibit modular forms, and these are provided by the Riemann–Roch theorem. For example, for $k = 2$, we have $\dim_{\mathbb{C}} M_2(\Gamma^1) = g = 0$ by (40.2.11). □

43.2.3. We are now in a position to prove an analogous statement to Theorem 40.3.11. Referring to Proposition (43.2.1), by part (a) we seek $a_1, a_2, a_2', a_3, a_3' \in \mathbb{Z}_{\geq 0}$ with

$$a_1 + \frac{a_2 + a_2'}{2} + \frac{a_3 + a_3'}{3} = \frac{k}{6}. \tag{43.2.4}$$

By part (b), we have $\dim_{\mathbb{C}} M_k(\Gamma^1) = 0$ for $k < 0$, and indeed, there are no solutions. For $k = 0$, there is a unique solution corresponding to the constant functions. For $k = 2$, there are no solutions, as follows. Let $f(z) \in M_2(\Gamma^1)$. Let γ_3 be a generator for the stabilizer at z_3. Then

$$f(z_3) = f(\gamma_3 z_3) = \jmath(\gamma_3; z_3)^2 f(z_3);$$

by the cocycle relation, we have $1 = \jmath(\gamma_3^3; z_3) = \jmath(\gamma_3; z_3)^3$ a nontrivial cube root of unity, so $f(z_3) = 0$ and $a_3 > 0$. Similarly $a_3' > 0$, and this contradicts (43.2.4).

Arguing in the same way, we find that the unique solution for $k = 4$ is $a_2 = a_2' = 0$ and $a_3 = a_3' = 1$; thus $M_4(\Gamma^1) = \mathbb{C}f_4$, and f_4 necessarily vanishes at z_2, z_2'. Similarly, for $k = 6$ we have only $a_3 = a_3' = 0$ and $a_2 = a_2' = 1$, with $M_6(\Gamma^1) = \mathbb{C}g_6$.

Continuing as in 43.2.3, we collect dimensions and spanning functions as in Table 43.2.5.

k	$\dim_{\mathbb{C}} M_k(\Gamma^1)$	Spanning functions
0	1	1
2	0	-
4	1	f_4
6	1	g_6
8	1	f_4^2
10	1	$f_4 g_6$
12	3	f_4^3, g_6^2, h_{12}
⋮	⋮	⋮
24	5	$f_4^6, f_4^3 g_6^2, f_4^3 h_{12}, g_6^4, g_6^2 h_{12}, h_{12}^2$

Table 43.2.5: Generators for $M(\Gamma)$

43.2.6. In weights 8, 10, we have products of forms seen previously. In weight $k = 12$, we find a third function $h_{12} \in M_{12}(\Gamma^1)$ spanning together with f_4^3, g_6^2. Continuing in

this way, finally in weight $k = 24$, we find 6 functions in a 5 dimensional space, and so they must satisfy an equation $r(f_4, g_6, h_{12}) \in \mathbb{C}[f_4, g_6, h_{12}]$, homogeneous of degree 24 if we give f_4, g_6, h_{12} the weights $4, 6, 12$.

Proposition 43.2.7. *We have*

$$M(\Gamma^1) \simeq \frac{\mathbb{C}[f_4, g_6, h_{12}]}{\langle r(f_4, g_6, h_{12})\rangle}.$$

Proof. The bound on the degrees of generators and relations in Theorem 43.9.6 makes this proposition immediate. It is also possible to give a proof with bare hands: see Exercise 43.9. ☐

43.2.8. We do not have Eisenstein series available in this setting, but the notion of taking averages 40.1.19 is still quite sensible: we find what are known as *Poincaré series*. Recall $j(\gamma; z) = cz + d$ for $\gamma = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in \mathrm{SL}_2(\mathbb{R})$. The square $j(\gamma; z)^2$ is well-defined on $\gamma \in \mathrm{PSL}_2(\mathbb{R})$.

For $k \in 2\mathbb{Z}_{\geq 2}$, we define the **Poincaré series**

$$P_k(z) = \sum_{\gamma \in \Gamma^1} j(\gamma; z)^{-k}.$$

Then $P_k(z)$ is nonzero, absolutely convergent on $\mathbf{H}^2$, uniformly on compact subsets, and satisfies

$$P_k(\gamma z) = j(\gamma; z)^k P_k(z). \tag{43.2.9}$$

The convergence statement is fairly tame, because the fundamental domain $\mathcal{D}$ is compact: it implies that the total integral

$$\int_{\mathcal{D}} \frac{(\operatorname{Im} z)^k}{|j(\gamma; z)|} \,\mathrm{d}\mu(z) < \infty$$

is finite, and the Poincaré series converges (absolutely) by comparison [Kat85, §1, Proposition 1]. The equality (43.2.9) follows from the cocycle relation (40.2.5). Therefore $P_k(z) \in S_k(\Gamma)$, and in particular we may take $f_4 = P_4$ and $g_6 = P_6$; with a bit more computation, one can also show that P_4^3, P_6^2, P_{12} are linearly independent, so that we may take $h_{12} = P_{12}$ as well.

43.2.10. A convenient and meaningful normalization of the functions above is given by Baba–Granath [BG2008, §3.1].

First, there are exactly two (necessarily optimal) embeddings $S = \mathbb{Z}[\sqrt{-6}] \hookrightarrow O$ by Example 30.7.5: we have $\# \mathrm{Cls}\, O = 1$ and $K(\sqrt{-6})$ is ramified at $p = 2, 3 \mid D = 6$, so $m(S, O; O^\times) = h(\mathbb{Z}[\sqrt{-6}]) = 2$. The fixed points of these two embeddings are distinct points $z_6, z_6' \in \mathcal{D}$. Explicitly, we note that

$$\mu = 3i + ij = -1 + 2i - j + 2k \tag{43.2.11}$$

has $\mu^2 + 6 = 0$, and choose its fixed point as z_6.

We rescale g_6 so that $f_4^3(z_6)/g_6^2(z_6) = \sqrt{-3}$, and we choose h_{12} such that $h_{12}(z_6) = h_{12}(z_6') = 0$, and rescale so that

$$r(f_4, g_6, h_{12}) = h_{12}^2 + 3g_6^4 + f_4^6 = 0. \tag{43.2.12}$$

Corollary 43.2.13. *The holomorphic map*

$$\begin{aligned} \Gamma^1(\mathcal{O}) &\to \mathbb{P}^2 \\ z &\mapsto (f_4^3(z) : g_6^2(z) : h_{12}(z)) \end{aligned} \tag{43.2.14}$$

has image the conic $X^1 : x^2 + 3y^2 + z^2 = 0$*, defining the canonical model over* $\mathbb{Q}$.

Proof. This result is attributed to Ihara by Kurihara [Kur79, Theorem 1-1(1)]; it is proven by Baba–Granath [BG2008, Theorem 3.10] along the lines above. □

We note that $X^1(\mathbb{R}) = \emptyset$; this is a general feature, see Proposition 43.7.2.

43.2.15. The Atkin–Lehner group

$$\mathrm{AL}(\mathcal{O}) = N_{B^\times}(\mathcal{O})/\mathbb{Q}^\times\mathcal{O}^\times \simeq (\mathbb{Z}/2\mathbb{Z})^2$$

has three nontrivial involutions w_2, w_3, w_6 with representatives having positive reduced norm. Explicitly, we have $w_6 = \mu$ by (43.2.11) and $w_2 = 1 + i$ and $w_3 = w_6/w_2 = 1 + i - j + k$. These involutions act on the space of modular forms as follows [BG2008, §3.1]:So for example $f_4(w_2 z) = -f_4(z)$.

	w_2	w_3	w_6
f_4	$-$	$-$	$+$
g_6	$+$	$-$	$-$
h_{12}	$-$	$+$	$-$

These involutions act on the canonical model X^1 by $w_2(x : y : z) = (x : -y : z)$, $w_3(-x : y : z)$, and $w_6(x : y : z) = (x : y : -z)$.

We choose a principal polarization (see Definition 43.6.4) on $\mathcal{O}$ by μ in (43.2.11). In this way, Main Theorem 43.1.4 provides that the curve X^1 parametrizes abelian surfaces with QM by $(\mathcal{O}, \mu)$.

43.2.16. The forgetful map $[(A_\tau, \iota_\tau)] \mapsto [A_\tau]$ which forgets the QM structure is the map [BG2008, Proposition 3.9]

$$\begin{aligned} j : X^1 &\to \mathbb{P}^1 \\ (x : y : z) &\mapsto \frac{16y^2}{9x^2} \end{aligned} \tag{43.2.17}$$

generically 4-to-1. The map j can fruitfully be thought of as an analogue of the classical elliptic j-invariant, mindful of the above technicalities: it parametrizes principally polarized complex abelian surfaces that *can be* equipped with a QM structure.

The Igusa invariants 43.3.5 of A_j where $j = j(\tau)$ are given by [BG2008, Proposition 3.6]

$$\begin{aligned}(I_2 : I_4 : I_6 : I_{10}) \\ &= (12(j+1) : 6(j^2+j+1) : 4(j^3-2j^2+1) : j^3) \\ &\in \mathbb{P}(2,4,6,10).\end{aligned} \tag{43.2.18}$$

There exists a genus 2 curve with these Igusa invariants if and only if $j = 0, -16/27$ or the Hilbert symbol

$$(-6j, -2(27j+16))_{\mathbb{Q}} = 1$$

is trivial.

Example 43.2.19. The two points with $j = 0, \infty$ are exactly those points which are not Jacobians of genus 2 curves: these correspond to points with CM by $\mathbb{Z}[\sqrt{-1}]$ and $\mathbb{Z}[\omega]$, and these abelian surfaces are the squares of the corresponding CM elliptic curves (with the product polarization). Elkies [Elk98, §3] computes equations and further CM points for discriminant 6.

Example 43.2.20. The case $j = -16/27$ corresponds to a CM point with discriminant $D = -24$ [BG2008, §3.3]: it is the Jacobian of the curve

$$y^2 = (1+\sqrt{2})x^6 - 3(7-3\sqrt{2})x^4 - 3(7+3\sqrt{2})x^2 + (1-\sqrt{2})$$

isomorphic to the product of the two elliptic curves with CM by $\mathbb{Z}[\sqrt{-6}]$ (but not with the product polarization).

Remark 43.2.21. For further reading to connect some of the dots above, see the article by Baba–Granath [BG2008], refining the work by Hashimoto–Murabayashi [HM95, Theorem 1.3] who give an explicit family of genus 2 curves whose Jacobians have QM by $\mathcal{O}$.

43.3 Genus 2 curves

We begin in the concrete setting of genus 2 curves. Let F be a perfect field with $\operatorname{char} F \neq 2$ and let F^{al} be an algebraic closure of F. Let X be a smooth projective curve of genus 2 over F.

43.3.1. Using Riemann–Roch in a manner analogous to the proof for elliptic curves (see e.g. Silverman [Sil2009, Proposition III.3.1(a)]), X is given by a **Weierstrass equation** of the form

$$y^2 = f(x) \tag{43.3.2}$$

where $f(x) \in F[x]$ is squarefree of degree 5 or 6. It follows that X is **hyperelliptic** over F, with map $x : X \to \mathbb{P}^1$ of degree 2.

If $(y')^2 = f'(x')$ is another Weierstrass equation for X, then it is related by a change of variables of the form

$$x' = \frac{ax+b}{cx+d}, \quad y' = \frac{ey}{(cx+d)^3} \tag{43.3.3}$$

with $ad - bc, e \in F^\times$. After such a change of variable, we may suppose without loss of generality that $\deg f = 6$.

Example 43.3.4. Let X^{al} be the base change of X to F^{al}. The automorphism group $\mathrm{Aut}(X^{\mathrm{al}})$ is a finite group containing the **hyperelliptic involution** $(x, y) \mapsto (x, -y)$. The possibilities for this group were classified by Bolza [Bol1887, p. 70]: when $\mathrm{char}\, F \neq 2, 3, 5$, the group $\mathrm{Aut}(X^{\mathrm{al}})$ is isomorphic to one of the groups

$$C_2, V_4, D_8, C_{10}, D_{12}, 2D_{12}, \widetilde{S_4}$$

of orders 2, 4, 8, 10, 12, 24, 48. A generic genus 2 curve over F^{al} has $\mathrm{Aut}(X^{\mathrm{al}}) \simeq C_2$.

43.3.5. We now seek invariants of the curve defined in terms of a model to classify isomorphism classes. We factor

$$f(x) = c \prod_{i=1}^{6} (x - a_i)$$

with $a_i \in F^{\mathrm{al}}$ the roots of f. We abbreviate $a_i - a_j$ by $(i\, j)$, and we define

$$\begin{aligned}
I_2 &:= (4c)^2 \sum (1\,2)^2(3\,4)^2(5\,6)^2, \\
I_4 &:= (4c)^4 \sum (1\,2)^2(2\,3)^2(3\,1)^2(4\,5)^2(5\,6)^2(6\,4)^2, \\
I_6 &:= (4c)^6 \sum (1\,2)^2(2\,3)^2(3\,1)^2(4\,5)^2(5\,6)^2(6\,4)^2(1\,4)^2(2\,5)^2(3\,6)^2, \\
I_{10} &:= (4c)^{10} \prod (1\,2)^2,
\end{aligned} \tag{43.3.6}$$

where each sum and product runs over the distinct expressions obtained by permuting the index set $\{1, \dots, 6\}$; by Galois theory, we have $I_2, I_4, I_6, I_{10} \in F$. In particular, we have

$$I_{10} = (4c)^{10} \prod_{1 \leq i < j \leq 6} (a_i - a_j)^2 = \mathrm{disc}(4f) \neq 0$$

is the discriminant of the polynomial $4f$. The invariants I_2, I_4, I_6, I_{10}, defined by Igusa [Igu60, p. 620] by modifying a set of invariants due to Clebsch, are known as the **Igusa–Clebsch invariants**.

Under a change of variable of the form (43.3.3), we have

$$(I_2', I_4', I_6', I_{10}') = (\delta^2 I_2, \delta^4 I_4, \delta^6 I_6, \delta^{10} I_{10})$$

where $\delta = e^2/(ad - bc)^3$. Accordingly, we define the **weighted** (2, 4, 6, 10)**-projective space**

$$\mathbb{P}(2, 4, 6, 10)(F^{\mathrm{al}}) := ((F^{\mathrm{al}})^4 \smallsetminus \{(0, 0, 0, 0)\})/\sim$$

under the equivalence relation

$$(I_2, I_4, I_6, I_{10}) \sim (\delta^2 I_2, \delta^4 I_4, \delta^6 I_6, \delta^{10} I_{10})$$

for all $\delta \in (F^{\mathrm{al}})^\times$; we write equivalence classes $(I_2 : I_4 : I_6 : I_{10}) \in \mathbb{P}(2, 4, 6, 10)(F^{\mathrm{al}})$.

Proposition 43.3.7. *The genus* 2 *curves* X *and* X' *over* F *are isomorphic over* F^{al} *if and only if*

$$(I_2 : I_4 : I_6 : I_{10}) = (I_2' : I_4' : I_6' : I_{10}') \in \mathbb{P}(2,4,6,10)(F^{\mathrm{al}}).$$

Proof. See Igusa [Igu60, Corollary, p. 632]. □

43.3.8. For arithmetic reasons (in particular to deal with problems in characteristic 2), Igusa [Igu60, pp. 617ff] defined the invariants [Igu60, pp. 621–622]

$$\begin{aligned} J_2 &:= I_2/8, \\ J_4 &:= (4J_2^2 - I_4)/96, \\ J_6 &:= (8J_2^3 - 160J_2J_4 - I_6)/576, \\ J_8 &:= (J_2J_6 - J_4^2)/4, \\ J_{10} &:= I_{10}/4096, \end{aligned} \tag{43.3.9}$$

now called the **Igusa invariants**, with $(J_2 : J_4 : J_6 : J_8 : J_{10}) \in \mathbb{P}(2,4,6,8,10)(F^{\mathrm{al}})$. Visibly, the Igusa–Clebsch invariants determine the Igusa invariants and vice versa.

Remark 43.3.10. One can also take ratios of these invariants with the same weight and define (three) absolute invariants analogous to the classical j-invariant of an elliptic curve, following Cardona–Nart–Pujolas [CNP2005] and Cardona–Quer [CQ2005].

Example 43.3.11. The locus of genus 2 curves with given automorphism group (cf. Example 43.3.4) can be described explicitly by the vanishing of polynomials in the Igusa(–Clebsch) invariants. For example, the unique genus 2 curve up to isomorphism over F^{al} with automorphism group C_{10} (when $\operatorname{char} F \neq 5$) is the curve defined by the equation $y^2 = x(x^5 - 1)$ with $(I_2 : I_4 : I_6 : I_{10}) = (0 : 0 : 0 : 1)$, with automorphism group generated by $(x, y) \mapsto (\zeta_5 x, -\zeta_5^3 y)$, where ζ_5 is a primitive fifth root of unity.

43.3.12. The group $\mathrm{Aut}_F(F^{\mathrm{al}})$ acts on $\mathbb{P}(2,4,6,10)(F^{\mathrm{al}})$ in each coordinate:

$$\sigma(I_2 : I_4 : I_6 : I_{10}) = (\sigma(I_2) : \sigma(I_4) : \sigma(I_6) : \sigma(I_{10}))$$

for $\sigma \in \mathrm{Aut}_F(F^{\mathrm{al}})$. Given a point $P \in \mathbb{P}(2,4,6,10)(F^{\mathrm{sep}})$, we define its **field of moduli** $M(P)$ to be the fixed field of F^{sep} under the stabilizer of P under this action. Just as in the case of ordinary projective space, the field $M(P)$ is the minimal field over which P is defined.

43.3.13. In this way, given a genus 2 curve, we have associated invariants of the curve that determine it up to isomorphism over F^{al}. We may also ask the inverse problem: given Igusa invariants $(J_k)_k$ with $J_{10} \neq 0$, find a genus 2 curve with the desired invariants. This problem has been solved explicitly by work of Mestre [Mes91] and Cardona–Quer [CQ2005].

We give a sketch of the generic case of curves whose only automorphism over F^{al} is the hyperelliptic involution, due to Mestre [Mes91]: in brief, the field of moduli may not be a field of definition for the desired genus 2 curve, but a quadratic extension will always suffice. Abbreviate $\mathbb{Q}[J] = \mathbb{Q}[J_2, J_4, J_6, J_8, J_{10}]$. First, Mestre constructs an explicit ternary quadratic form $L(J)$ and ternary cubic form $M(J)$ defined over $\mathbb{Q}[J]$.

Under substitution of generic invariants, the quadratic form $L(J)$ defines a quaternion algebra $B(J)$ over the field of moduli F of the point, and Mestre proves that there exists a curve X over a field $K \supseteq F$ with the desired Igusa invariants if and only if K is a splitting field for $B(J)$. The quaternion algebra $B(J)$ is accordingly called the **Mestre obstruction**. Over a field K where $B(J)$ splits, equivalently over a field K where the conic defined by $L(J) = 0$ has a K-rational point, we can parametrize $L(J)$ and by substituting into $M(J)$ we obtain a binary sextic form $f(x, z)$ with the property that $y^2 = f(x, 1)$ has the desired invariants.

43.4 Complex abelian varieties

Shifting gears, we pause to briefly recall some basic properties of complex abelian varieties, needed for our discussion of abelian surfaces. For further reference, see Birkenhake–Lange [BL2004], Mumford [Mum70], or Swinnerton–Dyer [Swi74].

Definition 43.4.1. A **complex torus** of dimension $g \in \mathbb{Z}_{\geq 1}$ is a complex manifold of the form $A = V/\Lambda$ where $g = \dim_{\mathbb{C}} V$ and $\Lambda \subseteq V$ is a lattice of rank $2g$. A **morphism** of complex tori $V/\Lambda \to V'/\Lambda'$ is a $\mathbb{C}$-linear map $\phi\colon V \to V'$ such that $\phi(\Lambda) \subseteq \Lambda'$.

Let $A = V/\Lambda$ be a complex torus of dimension g. Then $V \simeq \mathbb{C}^g$ and $\Lambda \simeq \mathbb{Z}^{2g}$ so $V/\Lambda \simeq (\mathbb{R}/\mathbb{Z})^{2g}$ as smooth real manifolds.

43.4.2. Suppose for concreteness (choosing a basis) that $V = \mathbb{C}^g$, working with column vectors. Choose a basis $\{\lambda_j\}_{j=1,\dots,2g}$ for Λ with $\lambda_j = (\lambda_{ij})_i^{\mathsf{t}} \in \mathbb{C}^g$. The matrix $\Pi = (\lambda_{ij})_{i,j} \in \mathrm{Mat}_{g\times 2g}(\mathbb{C})$ is called the **big period matrix** of the lattice Λ (with respect to the basis $\{\lambda_j\}_j$).

A change of basis of $\mathbb{C}^g$ corresponds to left multiplication by an element of $\mathrm{GL}_g(\mathbb{C})$ on Π and induces an isomorphism of complex tori. Writing

$$\Pi = \begin{pmatrix} P_1 & P_2 \end{pmatrix}, \qquad \text{with } P_1, P_2 \in \mathrm{GL}_g(\mathbb{C})$$

we have $P_2^{-1}\Pi = \begin{pmatrix} \Omega & 1 \end{pmatrix}$, and $\Omega = P_2^{-1}P_1 \in \mathrm{GL}_g(\mathbb{C})$. Therefore every complex torus is isomorphic to a torus of the form $\mathbb{C}^g/(\Omega\mathbb{Z}^g + \mathbb{Z}^g)$ for some $\Omega \in \mathrm{GL}_g(\mathbb{C})$, called the **small period matrix**.

Definition 43.4.3. A complex torus A is a **complex abelian variety** if there exists a holomorphic embedding $A \hookrightarrow \mathbb{P}^n(\mathbb{C})$ for some $n \geq 1$.

Remark 43.4.4. Every complex torus of dimension $g = 1$ is an abelian variety, indeed an elliptic curve, by the theory of classical Eisenstein series (see 40.1.11). But the case $g = 1$ is quite special! For a general lattice $\Lambda \subseteq \mathbb{C}^g$ with $g \geq 2$, there will probably be no meromorphic functions on $\mathbb{C}^g/\Lambda$ and in particular there will be no way to realize the torus as a projective algebraic variety.

The conditions under which a complex torus is a complex abelian variety are given by the following conditions, due to Riemann.

Definition 43.4.5. A matrix $\Pi \in \mathrm{Mat}_{g\times 2g}(\mathbb{C})$ is a **Riemann matrix** if there is a alternating matrix $E \in \mathrm{M}_{2g}(\mathbb{Z})_{\mathrm{alt}}$ with $\det E \neq 0$ such that:

(i) $\Pi E^{-1}\Pi^{\mathrm{t}} = 0$, and
(ii) $\sqrt{-1}\Pi E^{-1}\Pi^*$ is a positive definite Hermitian matrix, where $^* = {}^{-\mathrm{t}}$ denotes conjugate transpose.

Conditions (i) and (ii) are called the **Riemann relations**.

Theorem 43.4.6. *Let* $A = \mathbb{C}^g/(\Pi\mathbb{Z}^{2g})$ *be a complex torus with* $\Pi \in \mathrm{Mat}_{g\times 2g}(\mathbb{C})$. *Then* A *is a complex abelian variety if and only if* Π *is a Riemann matrix.*

Example 43.4.7. Let $f(x) \in \mathbb{C}[x]$ be a squarefree polynomial of degree $2g+1$ or $2g+2$ with $g \geq 1$. The equation $y^2 = f(x)$ defines an algebraic curve with projective closure X a nonsingular curve over $\mathbb{C}$. A basis of the holomorphic differential 1-forms on X is $\omega_i = x^{i-1}\mathrm{d}x/y$ for $i = 1,\dots,g$.

The set of points $X(\mathbb{C})$ has naturally the structure of a compact (connected) Riemann surface of genus $g \geq 1$. Let $\alpha_1,\beta_1,\dots,\alpha_g,\beta_g$ be a basis of the homology $H_1(X,\mathbb{Z})$ and suppose that this basis is **symplectic**: each closed loop α_i intersects β_i with (oriented) intersection number 1 and all other intersection numbers are 0, as in the following standard picture in Figure 43.4.8.

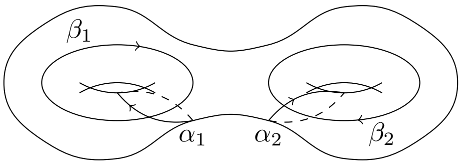


Figure 43.4.8: A standard surface with a symplectic homology basis

The integration pairing

$$\begin{aligned}\Omega^1 \times H_1(X,\mathbb{Z}) &\to \mathbb{C}\\ (\omega,\upsilon) &\mapsto \int_\upsilon \omega\end{aligned}$$

is nondegenerate, giving a map $H_1(X,\mathbb{Z}) \hookrightarrow \mathrm{Hom}_{\mathbb{C}}(\Omega^1,\mathbb{C})$. Let

$$\Lambda = \left\{ \left(\int_\upsilon \omega_i\right)^{\mathrm{t}} : \upsilon \in H_1(X,\mathbb{Z}) \right\} \subseteq \mathbb{C}^g.$$

A $\mathbb{Z}$-basis of Λ is given by the integrals with $\upsilon = \alpha_i,\beta_i$ for $i = 1,\dots,g$. Let

$$\mathrm{Jac}\,X := \mathrm{Hom}_{\mathbb{C}}(\Omega^1,\mathbb{C})/H_1(X,\mathbb{Z}) \simeq \mathbb{C}^g/\Lambda$$

be the **Jacobian** of X. Then $\mathrm{Jac}\,X$ is a complex torus of dimension g. It has big period matrix $\Pi = \begin{pmatrix} P_1 & P_2 \end{pmatrix}$, where

$$P_1 = \left(\int_{\alpha_i} \omega_j\right)_{i,j}, \quad P_2 = \left(\int_{\beta_i} \omega_j\right)_{i,j}.$$

By cutting open the Riemann surface along the given paths and applying Green's theorem, we verify that the big period matrix Π is indeed a Riemann matrix. Therefore the Jacobian $\mathrm{Jac}(X)$ is an abelian variety of genus g.

We now upgrade the above to a basis-free formulation.

Definition 43.4.9. Let

$$E\colon \Lambda \times \Lambda \to \mathbb{Z}$$

be an alternating $\mathbb{Z}$-bilinear map. Let $E_{\mathbb{R}}: V \times V \to \mathbb{R}$ is the scalar extension of E over $\mathbb{R}$ obtained from $\mathbb{R}\Lambda = V$.

We say E is a **Riemann form** for (V, Λ) if the following conditions hold:

(i) $E_{\mathbb{R}}(ix, iy) = E_{\mathbb{R}}(x, y)$ for all $x, y \in V$; and

(ii) The map

$$\begin{aligned} V \times V &\to \mathbb{R} \\ (x, y) &\mapsto E_{\mathbb{R}}(ix, y) \end{aligned}$$

defines a symmetric, positive definite $\mathbb{R}$-bilinear form on V.

43.4.10. Let E be a Riemann form for (V, Λ). If we choose a $\mathbb{Z}$-basis for Λ, we get a period matrix Π and a matrix for E which is a Riemann matrix (satisfying conditions (i)–(ii) of Definition 43.4.5), and conversely.

Proposition 43.4.11. *If E is a Riemann form for (V, Λ), then the map*

$$\begin{aligned} H\colon V \times V &\to \mathbb{C} \\ H(x, y) &= E_{\mathbb{R}}(ix, y) + iE_{\mathbb{R}}(x, y) \end{aligned} \tag{43.4.12}$$

is a positive definite Hermitian form on V with $\operatorname{Im} H = E_{\mathbb{R}}$.

Conversely, if H is a positive definite Hermitian form on V such that $\operatorname{Im} H(\Lambda) \subseteq \mathbb{Z}$, *then* $\operatorname{Im} H|_{\Lambda}$ *is a Riemann form for* (V, Λ).

Proof. This proposition can be checked directly, a bit of linear algebra fun: see Exercise 43.6. ∎

Example 43.4.13. For the torus $\mathbb{C}/(\mathbb{Z}i + \mathbb{Z})$, the forms

$$\begin{aligned} E(x_1 + ix_2, y_1 + iy_2) &= x_2 y_1 - x_1 y_2 \\ E_{\mathbb{R}}(i(x_1 + ix_2), y_1 + iy_2) &= x_1 y_1 - (-x_2 y_2) = x_1 y_1 + x_2 y_2 \\ H(x, y) &= x\overline{y}. \end{aligned}$$

define a Riemann form E, its associated (symmetric, positive definite) real part, and its associated (positive definite) Hermitian form H.

Definition 43.4.14. A complex torus $A = V/\Lambda$ equipped with a Riemann form is said to be **polarized**.

A homomorphism of polarized complex tori is a homomorphism $\phi\colon A \to A'$ of complex tori that respects the polarizations in the sense that the diagram

$$\begin{array}{ccc} \Lambda \times \Lambda & \xrightarrow{E} & \mathbb{Z} \\ {\scriptstyle (\phi,\phi)}\downarrow & \nearrow{\scriptstyle E'} & \\ \Lambda' \times \Lambda' & & \end{array}$$

commutes.

By Theorem 43.4.6 and Proposition 43.4.11, a polarized complex torus is an abelian variety, and accordingly we call it a polarized abelian variety.

We now seek to classify the possibilities for Riemann forms.

43.4.15. There is a normal form for alternating matrices, analogous to the Smith normal form of an integer matrix, called the **Frobenius normal form**. Let M be a free $\mathbb{Z}$-module of rank $2g$ equipped with an alternating form $E : M \times M \to \mathbb{Z}$. Then there exists a basis of M such that the matrix of E in this basis is

$$[E] = \begin{pmatrix} 0 & D \\ -D & 0 \end{pmatrix}$$

where $D = \operatorname{diag}(d_1, \dots, d_g)$ is a diagonal matrix with diagonal entries $d_i \in \mathbb{Z}_{\geq 0}$ and $d_1 \mid d_2 \mid \cdots \mid d_g$. The integers $d_1, \dots, d_g$ are uniquely determined by E, and are called the **elementary divisors** of E when all $d_i > 0$ (equivalently, E is nondegenerate).

Definition 43.4.16. A Riemann form E with elementary divisors $1, \dots, 1$ in its Frobenius normal is called a **principal** Riemann form.

Lemma 43.4.17. *Let $A = V/\Lambda$ be a polarized abelian variety, and suppose the Riemann form E has elementary divisors $d_1, \dots, d_g$. Then there is a basis for V and a basis for Λ such that the big period matrix of Λ is $\begin{pmatrix} \Omega & D \end{pmatrix}$ where $D = \operatorname{diag}(d_1, \dots, d_g)$, and Ω is symmetric and $\operatorname{Im} \Omega$ is positive definite.*

In particular, if A is principally polarized, then the conclusion of Lemma 43.4.17 holds for Ω, the small period matrix.

Proof. Compute the period matrix with respect to a basis in which the Riemann form is in Frobenius normal form. □

Example 43.4.18. Let $A_1 = V_1/\Lambda_1$ and $A_2 = V_2/\Lambda_2$ be two polarized abelian varieties, with Riemann forms E_1, E_2. Let $A = A_1 \times A_2 = V/\Lambda$, where $V = V_1 \oplus V_2$ and $\Lambda = \Lambda_1 \oplus \Lambda_2 \subseteq V_1 \oplus V_2 = V$. Then A can be equipped with the **product polarization** $E = E_1 \boxplus E_2$, defined by

$$E(x_1 + x_2, y_1 + y_2) = E_1(x_1, y_1) + E_2(x_2, y_2).$$

If E_1, E_2 are principal, then the product E is also principal.

43.4.19. Polarizations can be understood in terms of duality, as follows.

Let $A = V/\Lambda$ be a complex torus. A $\mathbb{C}$-**antilinear functional** on V is a function $f : V \to \mathbb{C}$ such that $f(x + x') = f(x) + f(x')$ for all $x, x' \in V$ and $f(ax) = \overline{a} f(x)$ for all $a \in \mathbb{C}$ and $x \in V$. Let $V^* = \operatorname{Hom}_{\overline{\mathbb{C}}}(V, \mathbb{C})$ be the $\mathbb{C}$-vector space of $\mathbb{C}$-antilinear functionals on V. Then V^* is a $\mathbb{C}$-vector space with $\dim_{\mathbb{C}} V = \dim_{\mathbb{C}} V^*$ and the underlying $\mathbb{R}$-vector space of V^* is canonically isomorphic to $\operatorname{Hom}_{\mathbb{R}}(V, \mathbb{R})$. The canonical $\mathbb{R}$-bilinear form

$$\begin{aligned} V^* \times V &\to \mathbb{R} \\ (f, x) &\mapsto \operatorname{Im} f(x) \end{aligned}$$

is nondegenerate, so

$$\Lambda^* := \{f \in V^* : \operatorname{Im} f(\Lambda) \subseteq \mathbb{Z}\}$$

is a lattice in V^*, called the **dual lattice** of Λ, and the quotient $A^\vee := V^*/\Lambda^*$ is a complex torus. Double antiduality and nondegeneracy gives a canonical identification $(V^*)^* \cong V$, giving a canonical identification $(A^\vee)^\vee \cong A$.

Now suppose A is polarized with E a Riemann form for (V, Λ), and let H be the associated Hermitian form (43.4.12). Then double duality induces a Riemann form E^* on (V^*, Λ^*), so $A^\vee$ is a polarized abelian variety. We have a $\mathbb{C}$-linear map

$$\begin{aligned} \lambda\colon V &\to V^* \\ x &\mapsto H(x,-) \end{aligned} \tag{43.4.20}$$

with the property that $\lambda(\Lambda) \subseteq \Lambda^*$. Since the form is nondegenerate, the induced homomorphism $\lambda\colon A \to A^\vee$ is an isogeny of polarized abelian varieties. The degree of the isogeny λ is equal to the product $d_1 \cdots d_g$ of the elementary divisors of E, so in particular if E is principal then λ is an isomorphism of principally polarized abelian varieties.

43.4.21. Let $A = V/\Lambda$ be a principally polarized complex abelian variety with Riemann form E. Let $\phi\colon A \xrightarrow{\sim} A^\vee$ be the isomorphism of principally polarized abelian varieties induced by (43.4.20). Then we define the **Rosati involution** associated to E (or ϕ) by

$$\begin{aligned} {}^\dagger\colon \operatorname{End}(A) &\to \operatorname{End}(A) \\ \alpha &\mapsto \alpha^\dagger = \phi^{-1}\alpha^\vee\phi \end{aligned} \tag{43.4.22}$$

where $\alpha^\vee\colon A^\vee \to A^\vee$ is the isogeny induced by pullback. The Rosati involution is uniquely defined by the condition

$$E(x, \alpha y) = E(\alpha^\dagger x, y) \tag{43.4.23}$$

for all $x, y \in \Lambda$.

Proposition 43.4.24. *The Rosati involution $\dagger$ is a positive involution on the $\mathbb{Q}$-algebra* $\operatorname{End}(A) \otimes \mathbb{Q}$.

Proof. Let $B := \operatorname{End}(A) \otimes \mathbb{Q}$, and let $\alpha \in B$ with $\alpha \neq 0$. Let $\beta := \alpha^\dagger\alpha$. Since $\beta^\dagger = \beta$, and † is the adjoint involution with respect to the positive definite Hermitian form H, the action of β on V is diagonalizable with positive real eigenvalues: indeed, if $x \in V$ is an eigenvector with eigenvalue λ, then

$$\lambda H(x,x) = H(\beta x, x) = H(\alpha x, \alpha x) \in \mathbb{R}_{>0}$$

and $H(x,x) \in \mathbb{R}_{>0}$, so $\lambda \in \mathbb{R}_{>0}$. The eigenvalues of β on $\operatorname{End}(A)$ by left multiplication are its eigenvalues with some multiplicity, and accordingly the trace $\operatorname{tr}(\beta) = \operatorname{tr}(\alpha^\dagger\alpha)$ is a nonempty sum of these eigenvalues, hence is positive. ∎

43.4.25. Following Lemma 43.4.17, we define the **Siegel upper-half space**

$$\mathfrak{H}_g = \{\tau \in \mathrm{M}_g(\mathbb{C}) : \tau^{\mathsf{t}} = \tau \text{ and } \operatorname{Im}\tau \text{ is positive definite}\}.$$

To $\tau \in \mathfrak{H}_g$, we associate the lattice $\Lambda_\tau = \tau\mathbb{Z}^g \oplus \mathbb{Z}^g \subset \mathbb{C}^g$ and the abelian variety $A_\tau = \mathbb{C}^g/\Lambda_\tau$ with principal polarization

$$E_\tau(\tau x_1 + x_2, \tau y_1 + y_2) := x_1^{\mathsf{t}} y_2 - x_2^{\mathsf{t}} y_1 = (x_1^{\mathsf{t}}, x_2^{\mathsf{t}}) J \begin{pmatrix} y_1 \\ y_2 \end{pmatrix}$$

where $J = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}$. By Lemma 43.4.17, every principally polarized complex abelian variety arises in this way.

Two elements $\tau, \tau' \in \mathfrak{H}_g$ give rise to isomorphic abelian varieties if and only if they arise from a symplectic change of basis of Λ if and only if they are in the same orbit under the group

$$\mathrm{Sp}_{2g}(\mathbb{Z}) = \{\gamma \in \mathrm{M}_{2g}(\mathbb{Z}) : \gamma^{\mathsf{t}} J \gamma = J\}$$

where $\mathrm{Sp}_{2g}(\mathbb{Z}) \circlearrowright \mathfrak{H}_g$ acts by

$$\tau \mapsto (a\tau + b)(c\tau + d)^{-1}, \qquad \text{for } \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in \mathrm{Sp}_{2g}(\mathbb{Z}).$$

These maps give a bijection between the set of principally polarized complex abelian varieties of dimension g and the quotient

$$\mathcal{A}_g(\mathbb{C}) := \mathrm{Sp}_{2g}(\mathbb{Z})\backslash\mathfrak{H}_g.$$

By the theory of theta functions, $\mathcal{A}_g(\mathbb{C})$ is the set of complex points of a quasi-projective variety defined over $\mathbb{Q}$ of dimension $(g^2 + g)/2$.

43.5 Complex abelian surfaces

We now specialize to the case $g = 2$ of principally polarized abelian surfaces; in this section, we describe their moduli and the relationship with genus 2 curves, in analogy with elliptic curves ($g = 1$).

We recall Example 43.4.7, where abelian varieties were obtained from Jacobians of curves—we now specialize this to the case $g = 2$. The following theorem links complex genus 2 curves, via their Jacobians, to complex abelian surfaces.

Theorem 43.5.1. *Let A be a principally polarized abelian surface over $\mathbb{C}$. Then exactly one of the two holds:*

(i) *$A \simeq \operatorname{Jac} X$ is isomorphic as a principally polarized abelian surface to the Jacobian of a genus 2 curve X equipped with its natural polarization; or*

(ii) *$A \simeq E_1 \times E_2$ is isomorphic as a principally polarized abelian surface to the product of two elliptic curves equipped with the product polarization.*

43.5.2. In case (i) of Theorem 43.5.1, we say that A is **indecomposable** (as a principally polarized abelian surface, up to isomorphism). It is possible for A to be indecomposable as a principally polarized abelian surface and yet A is not simple, so A is *isogenous* to the product of elliptic curves. In case (ii), we say A is **decomposable**, and this case arises if and only if there is a basis e_1, e_2 for $\mathbb{C}^2$ such that

$$\Lambda = \Lambda_1 e_1 \oplus \Lambda_2 e_2$$

where $\Lambda_1, \Lambda_2 \subseteq \mathbb{C}$.

We now pursue an explicit version of Theorem 43.5.1, linking the algebraic description (section 43.3) to the analytic description (section 43.4), in a manner analogous to the construction of Eisenstein series for elliptic curves ($g = 1$) in 40.1.11 and 40.1.19.

43.5.3. For brevity, let $\Gamma := \mathrm{Sp}_4(\mathbb{Z})$, let

$$\jmath(\gamma;\tau) = c\tau + d \in \mathrm{M}_2(\mathbb{C}), \quad \text{for } \gamma = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in \Gamma \text{ and } \tau \in \mathfrak{H}_2,$$

where $a, b, c, d \in \mathrm{M}_2(\mathbb{Z})$, and let

$$\Gamma_\infty = \left\{ \gamma = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in \Gamma : c = 0 \right\}.$$

We define for $k \in 2\mathbb{Z}_{\geq 2}$ the **Eisenstein series**

$$\begin{aligned} \psi_k &: \mathfrak{H}_2 \to \mathbb{C} \\ \psi_k(\tau) &:= \sum_{\Gamma_\infty \gamma \in \Gamma_\infty \backslash \Gamma} \det \jmath(\gamma;\tau)^{-k}. \end{aligned}$$

As for classical Eisenstein series, $\psi_k(\tau)$ is absolutely convergent on compact domains. By design, the function ψ_k has a natural invariance under Γ:

$$\psi_k(\gamma\tau) = (\det \jmath(\gamma;\tau))^k \psi_k(\tau) \tag{43.5.4}$$

for all $\gamma \in \Gamma$ and $\tau \in \mathfrak{H}_2$.

We define two further functions:

$$\begin{aligned} \chi_{10} &:= -\frac{43867}{2^{12}3^5 5^2 7^1 53^1}(\psi_4\psi_6 - \psi_{10}) \\ \chi_{12} &:= \frac{131 \cdot 593}{2^{13}3^7 5^3 7^2 337^1}(3^2 7^2 \psi_4^3 + 2^1 5^3 \psi_6^2 - 691\psi_{12}) \end{aligned}$$

(The constants are taken so that the Fourier expansion is appropriately normalized; their precise nature can be safely ignored on a first reading.)

The function χ_{10} is somewhat analogous to the classical function Δ, in the following sense.

Lemma 43.5.5. *Let $\tau \in \mathfrak{H}_2$. Then $\chi_{10}(\tau) = 0$ if and only if A_τ is decomposable (as a principally polarized abelian variety).*

In other words, the vanishing locus of χ_{10} is precisely where case (ii) of Theorem 43.5.1 holds and the abelian surface is not isomorphic to the Jacobian of a genus 2 curve (as a principally polarized abelian surface).

Remark 43.5.6. More generally, a **(classical) Siegel modular form** of weight $k \in 2\mathbb{Z}$ for the group $\Gamma = \mathrm{Sp}_4(\mathbb{Z})$ is a holomorphic function $f\colon \mathfrak{H}_2 \to \mathbb{C}$ such that

$$f(\gamma\tau) = (\det \jmath(\gamma;\tau))^k f(\tau)$$

for all $\gamma \in \mathrm{Sp}_4(\mathbb{Z})$ and $\tau \in \mathfrak{H}_2$. (By the Koecher principle, such a function is automatically holomorphic at infinity in a suitable sense, and so the conditions at cusps 40.2.12 for classical modular forms do not arise here.)

Let $M_k(\Gamma)$ be the $\mathbb{C}$-vector space of Siegel modular forms of weight k; then $M_k(\Gamma)$ is finite-dimensional, $M_k(\Gamma) = \{0\}$ for $k < 0$, and $M_0(\Gamma) = \mathbb{C}$ consists of constant functions. Let $M(\Gamma) = \bigoplus_{k \in 2\mathbb{Z}_{\geq 0}} M_k(\Gamma)$; then $M(\Gamma)$ has the structure of a graded $\mathbb{C}$-algebra under pointwise multiplication of functions. Igusa proved that

$$M(\Gamma) = \mathbb{C}[\psi_4, \psi_6, \chi_{10}, \chi_{12}]$$

in analogy with Theorem 40.3.11. Extending the analogy, Igusa also proved that $\psi_4, \psi_6, -4\chi_{10}, 12\chi_{12}$ have integer Fourier coefficients with content 1.

43.5.7. The Igusa–Clebsch invariants 43.3.5 can be expressed in terms of the functions above. The precise relationship was worked out by Igusa [Igu60, p. 620]: we have

$$\begin{aligned} I_2 &= -2^3 3^1 \frac{\chi_{12}}{\chi_{10}} \\ I_4 &= 2^2 \psi_4 \\ I_6 &= -\frac{2^3}{3}\psi_6 - 2^5 \frac{\psi_4 \chi_{12}}{\chi_{10}} \\ I_{10} &= -2^{14} \chi_{10} \end{aligned}$$

The functions $I_4, I_6, \chi_{10}, \chi_{12}$ are holomorphic, but I_2 is meromorphic. In other words, if X is a complex genus 2 curve with $\operatorname{Jac} X = A_\tau$ for $\tau \in \mathfrak{H}_2$, then the algebraic invariants of X can be computed in terms of the values of these functions evaluated at τ. This description is again analogous to the case of elliptic curves (cf. Remark 40.3.10).

Proposition 43.5.8. *Two indecomposable principally polarized abelian surfaces A_τ, $A_{\tau'}$ are isomorphic (as principally polarized abelian surfaces) if and only if*

$$(I_2 : I_4 : I_6 : I_{10})(\tau) = (I_2 : I_4 : I_6 : I_{10})(\tau') \in \mathbb{P}(2,4,6,10)(\mathbb{C}).$$

In other words, the Igusa(–Clebsch) invariants are naturally defined coordinates on the moduli space $\mathcal{A}_2(\mathbb{C})$ of abelian surfaces, a complex threefold by 43.4.25.

Proof. Combine 43.5.7 and Proposition 43.3.7. ∎

43.5.9. Let A be a principally polarized complex abelian surface. Let $\mathrm{End}(A)$ be the ring of endomorphisms of A, and let $B = \mathrm{End}(A) \otimes_{\mathbb{Z}} \mathbb{Q}$. By the classification theorem of Albert (Theorem 8.5.4), the $\mathbb{Q}$-algebra B is exactly one of the following:

(i) $B = \mathbb{Q}$, and we say A is **typical**;
(ii) $B = F$ a real quadratic field, and we say A has **real multiplication (RM)** by F;
(iii) B is an indefinite quaternion algebra over $\mathbb{Q}$, and we say A has **quaternionic multiplication (QM)** by B;
(iv) $B = K$ is a quartic CM field K, and we say A has **complex multiplication (CM)** by K; or
(v) $B = \mathrm{M}_2(K)$ where K is an imaginary quadratic field.

Each one of these 5 cases is interesting in its own right—but given the subject of this book, we concern ourselves essentially with only case (iii), where quaternion algebras play a defining role.

43.6 Abelian surfaces with QM

In this section, we consider moduli spaces of abelian surfaces with quaternionic multiplication. For further reference, see Lang [Lang82, §IX.4–5]. Throughout, let A be a principally polarized complex abelian surface with Riemann form E. Let B be an indefinite quaternion algebra over $\mathbb{Q}$ with $\operatorname{disc} B = D$, let

$$\iota_\infty : B \hookrightarrow B_{\mathbb{R}} \simeq \mathrm{M}_2(\mathbb{R}) \tag{43.6.1}$$

be a splitting over $\mathbb{R}$.

43.6.2. The Rosati involution † (defined in 43.4.21) is a positive involution on $\mathrm{End}(A)_{\mathbb{Q}}$. We classified positive involutions in section 8.4: specifically, when $\mathrm{End}(A)_{\mathbb{R}} \simeq \mathrm{M}_2(\mathbb{R})$, by Example 8.4.15 there exists $\mu \in \mathrm{End}(A)_{\mathbb{R}}^\times$ with $\mu^2 \in \mathbb{R}_{<0}$ such that

$$\alpha^\dagger = \mu^{-1}\overline{\alpha}\mu \tag{43.6.3}$$

for all $\alpha \in \mathrm{End}(A)$. The map † defines a $\mathbb{Q}$-antiautomorphism of $\mathrm{End}(A)$, so by the Skolem–Noether theorem, we must have $\mu \in \mathrm{End}(A)^\times$.

From now on, let O be a maximal order in B. (One can relax this hypothesis with some additional technical complications, but there is enough to wrangle with here already!) In light of 43.6.2 we make the following definition (cf. Rotger [Rot2004, §3]).

Definition 43.6.4. A **polarization** on O is an element $\mu \in O$ such that $\mu^2 \in \mathbb{Z}_{<0}$; a polarization is **principal** if $\mu^2 + D = 0$.

An isomorphism of polarized orders $(O, \mu) \simeq (O', \mu')$ is an isomorphism of orders $\phi \colon O \xrightarrow{\sim} O'$ such that $\phi(\mu) = \mu'$.

43.6.5. By Lemma 17.4.2 (employing the Skolem–Noether theorem), an isomorphism $\phi \colon O \xrightarrow{\sim} O'$ of rings is induced by conjugation by an element of $B^\times$. In particular, the polarized orders (O, μ) and (O, μ') are isomorphic as polarized orders if and only if there exists $\nu \in N_{B^\times}(O)$ such that $\mu' = \nu^{-1}\mu\nu$.

43.6.6. Every O has a principal polarization. Indeed, by the local-global principle for splitting/embeddings (Proposition 14.6.7), the field $K = \mathbb{Q}(\sqrt{-D})$ embeds in B because $K_p = \mathbb{Q}_p(\sqrt{-D})$ is a field for all $p \mid D$. Therefore there exists $\mu' \in B$ with $(\mu')^2 + D = 0$, and $\mu' \in O'$ for some maximal order O'. But by a consequence of strong approximation (Theorem 28.2.11), O is conjugate to O', so there exists a conjugate $\mu \in O$ with still $\mu^2 + D = 0$. (By the theory of optimal embeddings, this is also immediately implied by Example 30.7.5, which counts the number of $O^\times$-equivalence classes of optimal embeddings $\mathbb{Z}_K \hookrightarrow O$.)

Lemma 43.6.7. *Let μ be a principal polarization on O. Then $\mu \in DO^\sharp$, i.e.,* $\operatorname{trd}(\mu O) \subseteq D\mathbb{Z}$.

Proof. We may check the desired equality locally. If $p \nmid D$, then $\mu \in O_p^\times$ and $\operatorname{trd}(\mu O_p) = \operatorname{trd}(O_p) = \mathbb{Z}_p$. Otherwise, if $p \mid D$, then μ generates the normalizer group $N_{B_p^\times}(O_p)/(\mathbb{Q}_p^\times O_p^\times)$ by Exercise 23.4, and $\operatorname{trd}(\mu O_p) \subseteq p\mathbb{Z}_p$ as desired. ◻

For a principally polarized order (O, μ), we define the positive involution

$$\begin{aligned} {}^* : B &\to B \\ \alpha^* &= \mu^{-1}\overline{\alpha}\mu \end{aligned} \tag{43.6.8}$$

From now on, let (O, μ) be a principally polarized order.

Definition 43.6.9. A **quaternionic multiplication (QM) structure** by (O, μ) on A is an injective ring homomorphism $\iota\colon O \hookrightarrow \operatorname{End}(A)$ such that the induced homomorphism $\iota\colon B \hookrightarrow \operatorname{End}(A)_\mathbb{Q}$ respects involutions, i.e., the diagram

$$\begin{array}{ccc} B & \xrightarrow{\iota} & \operatorname{End}(A)_\mathbb{Q} \\ \downarrow{*} & & \downarrow{\dagger} \\ B & \xrightarrow{\iota} & \operatorname{End}(A)_\mathbb{Q} \end{array} \tag{43.6.10}$$

commutes, where † is the Rosati involution defined in (43.4.22).

We say A **has quaternionic multiplication (QM) by** (O, μ) if A can be equipped with a QM structure by (O, μ), and we say that A is a **QM abelian surface** if it has a QM structure for some (O, μ).

Writing out (43.6.10), for a QM structure ι we require that $\iota(\alpha)^\dagger = \iota(\alpha^*)$ for all $\alpha \in B$.

Definition 43.6.11. A **homomorphism** $(A, \iota) \to (A', \iota')$ of principally polarized complex abelian surfaces with QM by (O, μ) is a homomorphism $\phi\colon A \to A'$ of polarized abelian surfaces (respecting the polarization) that also respects ι, ι' in the sense that the diagram

$$\begin{array}{ccc} B & \xrightarrow{\iota'} & \operatorname{End}(A')_\mathbb{Q} \\ & \searrow{\iota} & \downarrow{\phi^*} \\ & & \operatorname{End}(A)_\mathbb{Q} \end{array}$$

commutes; an **isogeny** is a surjective homomorphism with finite kernel.

QM abelian surfaces can be constructed as follows.

43.6.12. Extend ι_∞ to a map $\iota_\infty : B \hookrightarrow B_{\mathbb{C}} \simeq \mathrm{M}_2(\mathbb{C})$. Let $\tau \in \mathbf{H}^2$. Let

$$\Lambda_\tau := \iota_\infty(O)\begin{pmatrix}\tau\\1\end{pmatrix} \subset \mathbb{C}^2.$$

Then Λ_τ is a lattice in $\mathbb{C}^2$, since $\mathrm{rk}_{\mathbb{Z}} O = 4$; let $A_\tau := \mathbb{C}^2/\Lambda_\tau$ be the associated complex torus. The map ι_∞ induces a natural injective ring homomorphism $\iota_\tau : O \to \mathrm{End}(A_\tau)$ since $\iota_\infty(O)\Lambda_\tau \subseteq \Lambda_\tau$ as O itself is closed under multiplication. Define the form

$$\begin{aligned} E_\tau : \Lambda_\tau \times \Lambda_\tau &\to \mathbb{Z} \\ \left(x\begin{pmatrix}\tau\\1\end{pmatrix}, y\begin{pmatrix}\tau\\1\end{pmatrix}\right) &\mapsto \frac{1}{D}\,\mathrm{trd}(\iota_\infty(\mu)x\overline{y}) \end{aligned} \tag{43.6.13}$$

with $x, y \in \iota_\infty(O)$. The form E_τ takes values in $\mathbb{Z}$ by Lemma 43.6.7.

The main result of this section is then the following theorem.

Main Theorem 43.6.14. *Let* $\Gamma = \iota_\infty(O^1)/\{\pm 1\} \subseteq \mathrm{PSL}_2(\mathbb{R})$. *Then the map*

$$\begin{aligned} \Gamma\backslash\mathbf{H}^2 &\leftrightarrow \left\{ \begin{array}{c} (A,\iota) \text{ principally polarized} \\ \text{complex abelian surfaces} \\ \text{with QM by } (O,\mu) \\ \text{up to isomorphism} \end{array} \right\} \\ \Gamma\tau &\mapsto [(A_\tau, \iota_\tau)] \end{aligned} \tag{43.6.15}$$

is a bijection.

The proof of this theorem occupies the rest of this section; it amounts to checking that various conditions and compatibilities are satisfied. The reader who is willing to take these as verified can profitably move along to the next section.

We begin by verifying the Riemann relations.

Lemma 43.6.16. *The form* $E = E_\tau$ *defined in (43.6.13) or its negative* $E = -E_\tau$ *is a Riemann form.*

This sign ambiguity was already present above, as the involution (43.6.8) is the same replacing μ by $-\mu$; working backwards with Lemma 43.6.16 in hand, we may fix the sign by replacing μ by $-\mu$ once and for all so that E_τ itself is a Riemann form for all τ.

Proof. E is alternating since

$$\mathrm{trd}(\iota_\infty(\mu)x\overline{x}) = \mathrm{nrd}(x)\,\mathrm{trd}(\mu) = 0$$

for all $x \in \iota_\infty(O)$.

Next, let

$$\varrho = \frac{1}{\operatorname{Im}\tau}\begin{pmatrix} \operatorname{Re}\tau & -|\tau|^2 \\ 1 & -\operatorname{Re}\tau \end{pmatrix} \in \mathrm{M}_2(\mathbb{R}).$$

Then

$$\det(\varrho) = \varrho\overline{\varrho} = \frac{-(\operatorname{Re}\tau)^2 + |\tau|^2}{(\operatorname{Im}\tau)^2} = 1$$

so $\varrho \in \mathrm{SL}_2(\mathbb{R})$, and moreover

$$\varrho\begin{pmatrix}\tau\\1\end{pmatrix} = \frac{1}{\operatorname{Im}\tau}\begin{pmatrix}\tau\operatorname{Re}\tau - |\tau|^2 \\ \tau - \operatorname{Re}\tau\end{pmatrix} = \begin{pmatrix} i\operatorname{Re}\tau - \operatorname{Im}\tau \\ i \end{pmatrix} = i\begin{pmatrix}\tau\\1\end{pmatrix}. \tag{43.6.17}$$

Therefore, for all $x, y \in \iota_\infty(O)$,

$$\begin{aligned} E\left(ix\begin{pmatrix}\tau\\1\end{pmatrix}, iy\begin{pmatrix}\tau\\1\end{pmatrix}\right) &= E\left(xi\begin{pmatrix}\tau\\1\end{pmatrix}, yi\begin{pmatrix}\tau\\1\end{pmatrix}\right) = E\left(x\varrho\begin{pmatrix}\tau\\1\end{pmatrix}, y\varrho\begin{pmatrix}\tau\\1\end{pmatrix}\right) \\ &= \operatorname{trd}(\iota_\infty(\mu)(x\varrho)\overline{y\varrho}) = \operatorname{trd}(\iota_\infty(\mu)x(\varrho\overline{\varrho})\,\overline{y}) \\ &= \operatorname{trd}(\iota_\infty(\mu)x\overline{y}) = E\left(x\begin{pmatrix}\tau\\1\end{pmatrix}, y\begin{pmatrix}\tau\\1\end{pmatrix}\right). \end{aligned} \tag{43.6.18}$$

We now show that $(x, y) \mapsto E_{\mathbb{R}}(ix, y)$ is a symmetric, positive definite $\mathbb{R}$-bilinear form on V. It is enough to verify this for $x, y \in \iota_\infty(O)$. In this calculation, to avoid clutter we write μ for $\iota_\infty(\mu)$. Following as above, first we show symmetry:

$$\begin{aligned} E\left(ix\begin{pmatrix}\tau\\1\end{pmatrix}, y\begin{pmatrix}\tau\\1\end{pmatrix}\right) &= \operatorname{trd}(\mu(x\varrho)\overline{y}) = \operatorname{trd}(y\overline{\varrho}\,\overline{x}\,\overline{\mu}) \\ &= \operatorname{trd}(\mu(y\varrho)\overline{x}) = E\left(iy\begin{pmatrix}\tau\\1\end{pmatrix}, x\begin{pmatrix}\tau\\1\end{pmatrix}\right) \end{aligned} \tag{43.6.19}$$

using that $\overline{\mu} = -\mu$ and $\overline{\rho} = -\rho$ since they have trace zero. For positivity, we replace ϱ with an expression in μ in order to simplify, and then apply positivity. Since $\mu^2 = -D$, if we let $\mu_1 = \mu/\sqrt{D}$ with $\sqrt{D} \in \mathbb{R}_{>0}$, then $\mu_1^2 = -1$. Since also $\varrho^2 = -1$, there exists $\delta \in \mathrm{GL}_2(\mathbb{R})$ such that $\delta^{-1}\mu_1\delta = \varrho$. From the calculation that

$$\overline{\mu_1} = \mu_1^{-1} = \sqrt{D}\mu^{-1},$$

we obtain

$$\overline{\varrho} = \overline{\delta}\,\overline{\mu_1}\,\overline{\delta^{-1}} = \overline{\delta}(\sqrt{D}\mu^{-1})\frac{\delta}{\operatorname{nrd}(\delta)} = \frac{\sqrt{D}}{\operatorname{nrd}(\delta)}\overline{\delta}\mu^{-1}\delta \tag{43.6.20}$$

and hence

$$\begin{aligned} E\left(ix\begin{pmatrix}\tau\\1\end{pmatrix}, x\begin{pmatrix}\tau\\1\end{pmatrix}\right) &= \operatorname{trd}(\mu(x\varrho)\overline{x}) = -\operatorname{trd}(\mu x\overline{\varrho}\,\overline{x}) \\ &= -\frac{\sqrt{D}}{\operatorname{nrd}(\delta)}\operatorname{trd}(\mu x\overline{\delta}\mu^{-1}\delta\overline{x}). \end{aligned} \tag{43.6.21}$$

This may look worse, but now we use positivity of * applied to $x\overline{\delta}$:

$$\operatorname{trd}\left((x\overline{\delta})(x\overline{\delta})^*\right) = \operatorname{trd}(x\overline{\delta}\mu^{-1}\delta\overline{x}\mu) = \operatorname{trd}(\mu x\overline{\delta}\mu^{-1}\delta\overline{x}) > 0.$$

It follows that $E\left(ix\begin{pmatrix}\tau\\1\end{pmatrix}, x\begin{pmatrix}\tau\\1\end{pmatrix}\right)$ is always either positive definite or negative definite (depending on the sign of $\operatorname{nrd}(\delta)$). ∎

Lemma 43.6.22. *The polarization induced by E is principal.*

Proof. Let $\phi\colon A \to A^\vee$ be the isogeny induced by E. Then the degree of ϕ is

$$\deg\phi = \det\left(E\left(x_i\begin{pmatrix}\tau\\1\end{pmatrix}, x_j\begin{pmatrix}\tau\\1\end{pmatrix}\right)\right)_{i,j} = \det(\operatorname{trd}((\iota_\infty(\mu)/D)x_i\overline{x_j}))_{i,j}$$

where x_i are a $\mathbb{Z}$-basis for $\iota_\infty(\Lambda)$. Thus

$$\begin{aligned}\det(\operatorname{trd}((\iota_\infty(\mu)/D)x_i\overline{x_j}))_{i,j} &= \frac{\operatorname{Nm}(\mu)}{D^4}\det(\operatorname{trd}(x_i\overline{x_j}))_{i,j}\\ &= \frac{\operatorname{nrd}(\mu)^2}{D^4}(\operatorname{discrd}O)^2 = \frac{1}{D^2}D^2 = 1.\end{aligned}$$

∎

Lemma 43.6.23. *The homomorphism $\iota_\tau\colon O \to \operatorname{End}(A_\tau)_{\mathbb{Q}}$ satisfies the compatibility (43.6.10), and E_τ is the unique compatible principal polarization on A_τ compatible with ι_τ.*

Proof. Abbreviate $\iota = \iota_\tau$. Let $\alpha \in O$. Then the Rosati involution is uniquely defined by the condition

$$E\left(x\begin{pmatrix}\tau\\1\end{pmatrix}, \iota(\alpha)y\begin{pmatrix}\tau\\1\end{pmatrix}\right) = E\left(\iota(\alpha)^\dagger x\begin{pmatrix}\tau\\1\end{pmatrix}, y\begin{pmatrix}\tau\\1\end{pmatrix}\right) \tag{43.6.24}$$

for all $x, y \in \iota_\infty(O)$, i.e.,

$$\operatorname{trd}(\iota_\infty(\mu)x\overline{\iota(\alpha)y}) = \operatorname{trd}(\iota_\infty(\mu)\iota(\alpha)^\dagger x\overline{y}). \tag{43.6.25}$$

Let $z = \iota(\alpha)$ and $x, y \in \iota_\infty(O)$. We verify that (43.6.25) holds by

$$\begin{aligned}\operatorname{trd}(\iota_\infty(\mu)z^\dagger x\overline{y}) &= \operatorname{trd}(\iota_\infty(\mu)(\iota_\infty(\mu)^{-1}\overline{z}\iota_\infty(\mu))x\overline{y})\\ &= \operatorname{trd}(\iota_\infty(\mu)x\overline{y}\,\overline{z}) = \operatorname{trd}(\iota_\infty(\mu)x\overline{zy})\end{aligned} \tag{43.6.26}$$

as desired.

To conclude, any other polarization corresponds to another positive involution of the form $\alpha \mapsto \nu^{-1}\overline{\alpha}\nu$ as in 43.6.2; scaling, we may take $\nu \in O$ such that $\operatorname{trd}(\nu O) \subset DO^\sharp$. In the proof of Lemma 43.6.22, we see that the degree of the polarization is equal to $\operatorname{nrd}(\nu)^2/D^2$, so it is principal if and only if $\operatorname{nrd}(\nu) = D$, i.e., $\nu^2 + D = 0$. But then by compatibility $\operatorname{trd}(\iota_\infty(\mu)x\overline{y}) = \operatorname{trd}(\iota_\infty(\nu)x\overline{y})$ for all $x, y \in O$, which implies $\mu = \nu$. ∎

Remark 43.6.27. Lemma 43.6.23 shows that one could be more relaxed in the definition of QM abelian surface in the following sense. Let A be a (not yet polarized) complex abelian surface, and let $\iota\colon O \hookrightarrow \operatorname{End}(A)$ be a ring homomorphism. Then there is a unique principal polarization on A such that the induced Rosati involution is compatible with μ in the sense of (43.6.10).

Proposition 43.6.28. *Every principally polarized complex abelian surface with QM by (O, μ) is isomorphic as such to one of the form (A_τ, ι_τ) for some $\tau \in \mathbf{H}^2$.*

Proof. Let (A, ι) be a principally polarized complex abelian surface with QM by (O, μ), and let $A = \mathbb{C}^2/\Lambda$. Then

$$\mathrm{End}(A) = \{\alpha \in \mathrm{M}_2(\mathbb{C}) : \alpha\Lambda \subseteq \Lambda\}.$$

Therefore $\iota\colon O \hookrightarrow \mathrm{End}(A)$ extends to an inclusion $B \hookrightarrow \mathrm{M}_2(\mathbb{C})$. By the Skolem–Noether theorem, every two inclusions are conjugate by an element of $\mathrm{GL}_2(\mathbb{C})$, acting by an isomorphism of A; so without loss of generality, we may suppose that ι extends to ι_∞.

We claim that $\Lambda = \iota_\infty(O)x$ for some $x \in \mathbb{C}^2$. Indeed, $\Lambda \otimes \mathbb{Q}$ has the structure of a left B-module with the same dimension as B as a $\mathbb{Q}$-vector space; by Exercise 7.6, it follows that $\Lambda \otimes \mathbb{Q} = Bx$ is free as a left B-module with $x \in \Lambda \subseteq \mathbb{C}^2$; thus $\Lambda = Ix$ where $I \subseteq B$ is a left O-ideal. Since O is maximal and therefore hereditary, I is invertible as a left O-ideal, and in particular I is sated. Now comes strong approximation: by Corollary 28.5.17, since B is indefinite and $\mathrm{Cl}^+ \mathbb{Z}$ is trivial, we conclude that I is principal, and therefore we can rewrite $\Lambda = \iota_\infty(O)x$ with $x \in \mathbb{C}^2$. By Lemma 43.4.17, we may suppose that $x = \begin{pmatrix} \tau & 1 \end{pmatrix}$ with $\operatorname{Im} \tau > 0$, so $\tau \in \mathbf{H}^2$.

Finally, the polarization agrees by the uniqueness in Lemma 43.6.23. ∎

We are now ready to finish the proof of Main Theorem 43.6.14.

Proof of Main Theorem 43.6.14. By Lemmas 43.6.16, 43.6.22, and 43.6.23, the association $\tau \mapsto (A_\tau, \iota_\tau)$ yields a principally polarized complex abelian surface with QM by (O, μ). By Proposition 43.6.28, the map is surjective.

Next, we show the map is well-defined up to the action of Γ. Let $\gamma = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in \Gamma$ and $\tau' = \gamma\tau$. Then

$$\begin{aligned} \Lambda_{\tau'} = \iota_\infty(O)\begin{pmatrix} \gamma\tau \\ 1 \end{pmatrix} &= \iota_\infty(O)(c\tau + d)^{-1}\begin{pmatrix} a\tau + b \\ c\tau + d \end{pmatrix} \\ &= (c\tau + d)^{-1}\iota_\infty(O)\gamma\begin{pmatrix} \tau \\ 1 \end{pmatrix} = (c\tau + d)^{-1}\Lambda_\tau. \end{aligned} \tag{43.6.29}$$

Therefore scalar multiplication by $(c\tau + d)^{-1}$ induces an isomorphism $A_\tau \to A_{\gamma\tau}$ of abelian surfaces; this map preserves the polarization by writing

$$(c\tau + d)x\begin{pmatrix} \tau' \\ 1 \end{pmatrix} = x\gamma\begin{pmatrix} \tau \\ 1 \end{pmatrix}$$

for $x \in \iota_\infty(O)$, and then verifying that

$$\mathrm{trd}(\mu x\overline{y}) = \mathrm{trd}(\mu(x\gamma)(\overline{y\gamma})) = \mathrm{trd}(\mu x\gamma\overline{\gamma}\,\overline{y}) = \mathrm{trd}(\mu x\overline{y}).$$

The induced map $\mathrm{End}(A_{\gamma\tau}) \to \mathrm{End}(A_\tau)$ obtained from conjugation by a scalar matrix is the identity, and the compatibility for the QM is then verified by the fact that for $\alpha \in B$,

$$\iota'(\alpha) = \iota_\infty(\alpha) = \iota(\alpha).$$

To conclude, suppose that $(A_\tau, \iota_\tau) \simeq (A_{\tau'}, \iota_{\tau'})$ with $\tau, \tau' \in \mathbf{H}^{2\pm}$; then there exists $\phi \in \mathrm{GL}_2(\mathbb{C})$ such that $\phi\Lambda_\tau = \Lambda_{\tau'}$ and ϕ commutes with $\iota_\infty(\alpha)$. Since $\iota_\infty(\alpha) \otimes \mathbb{C} = \mathrm{M}_2(\mathbb{C})$, we conclude that ϕ is central in $\mathrm{M}_2(\mathbb{C})$ and hence scalar. From

$$\phi\begin{pmatrix}\tau\\1\end{pmatrix} = \gamma\begin{pmatrix}\tau'\\1\end{pmatrix}$$

we conclude $\gamma \in \Gamma$ and then that $\phi = c\tau' + d$ so $\tau = \gamma\tau'$. □

Remark 43.6.30. Analogous to the case $\mathrm{SL}_2(\mathbb{Z})$, one may similarly define congruence subgroups of Γ^1; the objects then parametrized are QM abelian surfaces equipped with a subgroup of torsion points.

Remark 43.6.31. The "forgetful" map which forgets the QM structure ι gives a map of moduli from $\Gamma\backslash\mathbf{H}^2$ to $\mathcal{A}_2(C)$, but this map is not injective: it factors via the quotient by the larger group $\Gamma\langle\mu\rangle$. In other words, the bijection of Main Theorem 43.6.14 induces a bijection between $\Gamma\langle\mu\rangle\backslash\mathbf{H}^2$ and the set of isomorphism classes of principally polarized abelian surfaces A such that A has QM by (O, μ): accordingly, generically there will be two choices of QM structure on a surface A that can be given QM by (O, μ).

Remark 43.6.32. The results above for $F = \mathbb{Q}$ extend, but not in a straightforward way, to totally real fields F of larger degree $n = [F : \mathbb{Q}]$. If A is an abelian variety with $\dim(A) = g$ such that A has QM by B over F, then $4n \mid 2g$, so we must consider abelian varieties of dimension at least $2n$. If equality $g = 2n$ holds, then A is simple, and by Albert's classification of the possible endomorphism algebras of an abelian variety, B is either totally definite or totally indefinite. So if $t = 1$, then $F = \mathbb{Q}$ and B is totally indefinite.

Consequently, we must consider abelian varieties of larger dimension. The basic construction works as follows. First, one chooses an element $\mu \in O$ such that $\mu^2 \in \mathbb{Z}_F$ is totally negative. (If $\mathbb{Z}_F$ has strict class number 1, then one has $\operatorname{disc} B = \mathfrak{D} = D\mathbb{Z}_F$ with D totally positive and one can choose μ satisfying $\mu^2 = -D$.) Let $K = F(\sqrt{-D})$; note that since $K \hookrightarrow B$ we have $B_K = B \otimes_F K \cong M_2(K)$. Then the complex space $X^B(1)_\mathbb{C}$ parametrizes complex abelian $4n$-folds with endomorphisms (QM) by B_K and equipped with a particular action on F on the complex differentials of B. Amazingly, this moduli interpretation does not depend on the choice of K; but because of this choice, it is not canonical as for the case $F = \mathbb{Q}$.

43.7 Real points, CM points

Let $X^1 = \Gamma\backslash\mathbf{H}^2$. Then X^1 has the structure of a complex 1-orbifold. We conclude this chapter with some discussion about real structures.

43.7.1. By Example 28.6.5, there exists $\epsilon \in O^\times$ such that $\operatorname{nrd}(\epsilon) = -1$. Then $\epsilon^2 \in O^1$ and ϵ normalizes O^1, so the action of ϵ (as in (33.3.12)) defines an anti-holomophic involution on $X(\Gamma)$ that is independent of the choice of ϵ: this gives the natural action of complex conjugation on $X(\Gamma)$.

With respect to this real structure, and in view of Main Theorem 43.6.14, we may ask if there are any principally polarized abelian surfaces with QM by (O, μ) with both the surface and the QM defined over $\mathbb{R}$. When $B \simeq \mathrm{M}_2(\mathbb{Q})$, then the element $\epsilon = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}$ acts by complex conjugation, and the real points are those points on the imaginary axis (the points with real j-invariant). More generally, the answer is provided by the following special case of a theorem of Shimura [Shi75, Theorem 0].

Proposition 43.7.2 (Shimura). *If B is a division algebra, then $X^1(\mathbb{R}) = \emptyset$.*

Proof. We follow Ogg [Ogg83, §3]. Let $\iota_\infty(\epsilon) = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$. If $X^1(\mathbb{R}) \neq \emptyset$, then by 43.7.1 there exists $z \in \mathbf{H}^2$ such that

$$z = \iota_\infty(\epsilon) \cdot z = \frac{a\bar{z} + b}{c\bar{z} + d}.$$

Then $a\bar{z} + b = c|z|^2 + dz$; taking imaginary parts we find $-a \operatorname{Im} z = d \operatorname{Im} z$ so $a + d = \operatorname{trd} \epsilon = 0$ and $\epsilon^2 - 1 = 0$. Since B is a division algebra, we conclude $\epsilon = \pm 1$, a contradiction. □

It may nevertheless happen that certain *quotients* of X^1 by Atkin–Lehner involutions may have real points.

Remark 43.7.3. A similar issue arises for CM elliptic curves: such a curve, together with all its endomorphisms, cannot be defined over $\mathbb{R}$.

Just as $Y(1) = \mathrm{SL}_2(\mathbb{Z})\backslash\mathbf{H}^2$ parametrizes isomorphism classes of elliptic curves, among them are countably many elliptic curves whose endomorphism algebra is larger than $\mathbb{Z}$: these are the elliptic curves with complex multiplication, and they correspond to points in $\mathbf{H}^2$ that are quadratic irrationalities, so $\mathbb{Q}(\tau) = K$ is an imaginary quadratic field and $S = \mathbb{Z}[\tau] \subseteq K$ is an imaginary quadratic order. By the theory of complex multiplication, the corresponding j-invariants are defined over the ring class field $H \supseteq K$ with $\mathrm{Gal}(H \,|\, K) \simeq \mathrm{Pic}(S)$, and there is an explicit action of $\mathrm{Gal}(H \,|\, K)$ on this set.

In a similar way, on X^1 we have CM points, given by complex abelian surfaces with extra endomorphisms, defined more precisely as follows.

43.7.4. Let $K \supseteq \mathbb{Q}$ be an imaginary quadratic field and suppose that $\iota_K \colon K \hookrightarrow B$ embeds. Let $S = K \cap O$, so that $S \hookrightarrow O$ is optimally embedded. Suppose the image of this embedding is given by $S = \mathbb{Z}[\nu]$ where $\nu \in O$. Let $\tau = \tau_\nu$ be the unique fixed point of $\iota_\infty(\nu)$ in $\mathbf{H}^2$; a point of $\mathbf{H}^2$ that arises this way is called a **CM point**.

The corresponding abelian surface A_τ visibly has $\mathrm{M}_2(S) \hookrightarrow \mathrm{End}(A_\tau)$, and in particular $\mathrm{End}(A_\tau)_{\mathbb{Q}} \simeq \mathrm{M}_2(K)$ as this is as large as possible.

43.8 ∗ Canonical models

In this section, we sketch the theory of canonical models for the curves X^1.

Theorem 43.8.1 (Shimura [Shi67, Main Theorem I (3.2)]). *There exists a projective, nonsingular curve $X^1_{\mathbb{Q}}$ defined over $\mathbb{Q}$ and a holomorphic map*

$$\varphi\colon \mathbf{H}^2 \to X^1_{\mathbb{Q}}(\mathbb{C})$$

that induces an analytic isomorphism

$$\varphi\colon \Gamma^1\backslash\mathbf{H}^2 \to X^1_{\mathbb{Q}}(\mathbb{C}).$$

43.8.2. The curve $X^1_{\mathbb{Q}}$ is made canonical (unique up to isomorphism) according to the field of definition of CM points (see 43.7.4).

Let $z \in \mathbf{H}^2$ be a CM point with S of discriminant D. Let $H \supseteq K$ be the ring class extension $H \supseteq K$ with $\mathrm{Gal}(H\,|\,K) \simeq \mathrm{Pic}(S)$. Then $\phi(z) \in X^1_{\mathbb{Q}}(H)$. Moreover, there is an explicit law, known as the *Shimura reciprocity law*, which describes the action of $\mathrm{Gal}(H\,|\,K)$ on them: to a class $[\mathfrak{c}] \in \mathrm{Pic}\, S$, we have

$$\iota_K(\mathfrak{c})O = \xi O \tag{43.8.3}$$

for some $\xi \in O$, and if $\sigma = \mathrm{Frob}_{\mathfrak{c}} \in \mathrm{Gal}(H\,|\,K)$ under the Artin isomorphism, then

$$\sigma(\phi(z)) = \phi(\xi^{-1}z). \tag{43.8.4}$$

For more detail, see Shimura [Shi67, p. 59]; and for an explicit, algorithmic version, see Voight [Voi2006].

43.8.5. Before continuing, we link back to the idelic, double-coset point of view, motivated in section 38.6 and given in general in section 38.7.

The difference between a lattice in $\mathbb{R}^2$ and a lattice in $\mathbb{C}$ is an identification of $\mathbb{R}^2$ with $\mathbb{C}$, i.e., an injective $\mathbb{R}$-algebra homomorphism $\psi\colon \mathbb{C} \to \mathrm{End}_{\mathbb{R}}(\mathbb{R}^2)$; we call ψ a **complex structure**.

There is a bijection between the set of complex structures and the set $\mathbb{C}\smallsetminus\mathbb{R} = \mathbf{H}^{2\pm}$ as follows. A complex structure $\psi\colon \mathbb{C} \to \mathrm{End}_{\mathbb{R}}(\mathbb{R}^2)$, by $\mathbb{R}$-linearity is equivalently given by the matrix $\psi(i) \in \mathrm{GL}_2(\mathbb{R})$ satisfying $\psi(i)^2 = -1$. By rational canonical form, every such matrix is similar to $\begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}$, i.e., there exists $\beta \in \mathrm{GL}_2(\mathbb{R})$ such that

$$\psi(i) = \beta^{-1}\begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}\beta,$$

and β is well-defined up to the centralizer of $\begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}$; this matrix acts by fixing $i \in \mathbf{H}^2$, and it follows that this centralizer is precisely $\mathbb{R}^\times\,\mathrm{SO}(2)$. Finally, we have $\mathrm{GL}_2(\mathbb{R})/\mathbb{R}^\times\,\mathrm{SO}(2) \xrightarrow{\sim} \mathbf{H}^{2\pm}$ under $\beta \mapsto \beta i$.

In this way,

$$\begin{aligned} X^1 = \Gamma^1\backslash\mathbf{H}^2 &\leftrightarrow O^1\backslash\mathbf{H}^{2\pm} \\ &\leftrightarrow O^1\backslash(\mathrm{GL}_2(\mathbb{R})/\mathbb{R}^\times\,\mathrm{SO}(2)) \\ &\leftrightarrow B^\times\backslash(\widehat{B}^\times/\widehat{O}^\times \times \mathrm{GL}_2(\mathbb{R})/\mathbb{R}^\times\,\mathrm{SO}(2)). \end{aligned} \tag{43.8.6}$$

(The final line is just an expression of the fact that $B^\times\backslash\widehat{B}^\times/\widehat{O}^\times$ is a set with one element, by strong approximation; it is placed there for comparison with other settings, where class numbers may add spice.) So the bijection (43.8.6) says that X^1 parametrizes O-lattices in B with a complex structure up to homothety. In the previous few sections, we showed how such lattices equipped with complex structure can be interpreted as a moduli space for abelian surfaces with quaternion multiplication.

We conclude this section with some more parting comments on representability.

Remark 43.8.7. Let $\mathsf{Sch}_{\mathbb{Q}}$ denote the category of schemes over $\mathbb{Q}$ under morphisms of schemes, and let Set denote the category of sets under all maps of sets. Let $\mathcal{F} : \mathsf{Sch}_{\mathbb{Q}} \to \mathsf{Set}$ be a contravariant functor. Then $X \in \mathsf{Sch}_{\mathbb{Q}}$ is a *coarse moduli space* for $\mathcal{F}$ (or X *coarsely represents* $\mathcal{F}$) if there exists a natural transformation $\Phi\colon \mathcal{F}(-) \to \operatorname{Hom}(-, X)$ which satisfies:

(i) $\Phi\colon \mathcal{F}(k) \xrightarrow{\sim} \operatorname{Hom}(k, X)$ is bijective if k is algebraically closed (and $\operatorname{char} k = 0$); and

(ii) Φ is *universal*: if (Z, Ψ) is another such coarse moduli space, then there is a unique commutative diagram

$$\begin{array}{ccc} \mathcal{F}(-) & \longrightarrow & \operatorname{Hom}(-, Z) \\ & \searrow & \Big\downarrow \exists! \\ & & \operatorname{Hom}(-, X) \end{array}$$

By Yoneda's lemma, condition (ii) is equivalent to a unique (commuting) morphism $X \to Z$.

We then define a functor $\mathcal{F}_O : \mathsf{Sch}_{\mathbb{Q}} \to \mathsf{Set}$ which associates to S the set of isomorphism classes of abelian schemes A over S—which can be thought of families of abelian surfaces parametrized by S—together with a map $\iota\colon O \hookrightarrow \operatorname{End}_S(A)$.

Deligne [Del71] has shown that the functor $\mathcal{F}_O$ is coarsely representable by a curve $X^1_{\mathbb{Q}}$ over $\mathbb{Q}$. By uniqueness and the solution to the moduli problem over $\mathbb{C}$, there is a map $\Gamma^B(1)\backslash\mathbf{H}^2 \xrightarrow{\sim} X^1_{\mathbb{Q}}(\mathbb{C})$ which is in fact an analytic isomorphism. Together with the field of definition of CM points, we recover the canonical model (Theorem 43.8.1).

43.9 ∗ Modular forms

In this final section, we sketch aspects of the theory of modular forms for arithmetic Fuchsian groups.

We restore a bit more generality: let F be a totally real field of degree $n = [F : \mathbb{Q}]$, let B be a quaternion algebra over F that split at exactly one real place corresponding to an embedding $\iota_\infty : B \hookrightarrow \mathrm{M}_2(\mathbb{R})$, let $O \subset B$ be an order, and let $\Gamma \leq \mathrm{PSL}_2(\mathbb{R})$ be a group commensurable with $\Gamma^1(O) = \iota_\infty(O^1)/\{\pm 1\}$.

Let $Y = \Gamma\backslash\mathbf{H}^2$, let $X = \Gamma\backslash\mathbf{H}^{2(*)}$ be its completion, and call the set $X \smallsetminus Y$ the set of **cusps**. We recall the notion of orbifold from section 34.8. Then X has the structure of a good complex 1-orbifold with signature $(g; e_1, \dots, e_r; \delta)$: X has genus g, there are exactly r elliptic points P_i of orders e_i, and δ cusps $Q_1, \dots, Q_\delta$. The hyperbolic area of

X is written $\mu(X)$, and can be computed as the area of a suitable fundamental domain.

As in the introduction 43.1.6, we make the following definition.

Definition 43.9.1. Let $k \in 2\mathbb{Z}_{\geq 0}$. A map $f\colon \mathbf{H}^2 \to \mathbb{C}$ is **weight k-invariant** under Γ if

$$f(\gamma z) = (cz+d)^k f(z) \quad \text{for all } \gamma = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in \Gamma. \tag{43.9.2}$$

A **modular form for Γ of weight k** is a holomorphic function $f\colon \mathbf{H}^2 \to \mathbb{C}$ that is weight k invariant and is holomorphic at the cusps.

Let $M_k(\Gamma)$ be the $\mathbb{C}$-vector space of modular forms for Γ and let

$$M(\Gamma) := \bigoplus_{k \in 2\mathbb{Z}_{\geq 0}} M_k(\Gamma); \tag{43.9.3}$$

then $M(\Gamma)$ is a graded $\mathbb{C}$-algebra under multiplication.

We can understand the degree of divisors of $M_k(\Gamma)$ as follows.

Theorem 43.9.4. *For $f \in M_k(\Gamma)$, we have*

$$\sum_{\Gamma z \in \Gamma\backslash \mathbf{H}^2} \frac{1}{\#\operatorname{Stab}_\Gamma(z)} \operatorname{ord}_z(f) = \frac{k}{4\pi}\mu(X).$$

Proof. See Shimura [Shi71, Proposition 2.16, Theorem 2.20]. □

By an application of the theorem of Riemann–Roch and the description of modular forms behind Theorem 43.9.4, we find that $\dim_{\mathbb{C}} M_k(\Gamma)$ can be expressed in terms of k and the signature of Γ as follows.

Theorem 43.9.5. *We have*

$$\dim_{\mathbb{C}} M_k(\Gamma) = \begin{cases} 1, & \text{if } k = 0; \\ g + \max(0, \delta - 1), & \text{if } k = 2; \\ (k-1)(g-1) + \dfrac{k}{2}\delta + \displaystyle\sum_{i=1}^{r} \left\lfloor \frac{k}{2}\left(1 - \frac{1}{e_i}\right)\right\rfloor, & \text{if } k > 2. \end{cases}$$

Proof. See Shimura [Shi71, Theorem 2.23]. □

The above formulas can be proven in a different and straightforward way in the language of *stacky curves*. This description gives the following further information on $M(\Gamma)$.

Theorem 43.9.6 (Voight–Zureick-Brown [VZB2015]). *Let $e = \max(1, e_1, \dots, e_r)$. Then the ring $M(\Gamma)$ is generated as a $\mathbb{C}$-algebra by elements of weight at most $6e$ with relations in weight at most $12e$.*

(The case of signature $(0; 2, 2, 3, 3; 0)$ from section 43.2 is described [VZB2015, Table (IVa-3)] as a weighted plane curve of degree 12 in $\mathbb{P}(6, 3, 2)$.)

Remark 43.9.7. An appealing mechanism for working explicitly with modular forms in the absence of cusps is provided by power series expansions: for an introduction with computational aspects, see Voight–Willis [VW2014] and the references therein.

Exercises

1. Let $g = 1$ and consider a period matrix $\Pi = \begin{pmatrix}\omega_1 & \omega_2\end{pmatrix}$ with $\omega_1, \omega_2 \in \mathbb{C}$. Let $E = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}$. Show that in Definition 43.4.5 for E that the condition (i) is automatic and condition (ii) is equivalent to $\operatorname{Im}(\omega_2/\omega_1) > 0$.
2. Let $\Pi \in \operatorname{Mat}_{g\times 2g}(\mathbb{C})$. Show that Π is a period matrix for a complex torus if and only if $\begin{pmatrix}\Pi \\ \overline{\Pi}\end{pmatrix}$, the matrix obtained by vertically stacking Π on top of its complex conjugate $\overline{\Pi}$, is nonsingular.
3. Let $E\colon \Lambda \times \Lambda \to \mathbb{Z}$ be a $\mathbb{Z}$-bilinear form that satisfies conditions (i) and (ii) of Definition 43.4.9 (so a Riemann form but without the condition that E is alternating). Show that $E_{\mathbb{R}}$ (and E) are alternating.
4. Show that the symmetry (43.6.19) implies the equality
$$E\left(ix\begin{pmatrix}\tau\\1\end{pmatrix}, iy\begin{pmatrix}\tau\\1\end{pmatrix}\right) = E\left(x\begin{pmatrix}\tau\\1\end{pmatrix}, y\begin{pmatrix}\tau\\1\end{pmatrix}\right)$$
from (43.6.18) directly using Exercise 43.3 (without ϱ).

▹5. Let $A = V/\Lambda$ be a complex abelian variety.

(a) Let $\phi\colon A \to A'$ be an isogeny with $n = \#\ker(\phi)$. Show that there exists a unique isogeny $\psi\colon A' \to A$ such that $\psi \circ \phi = n_A$ is multiplication by n on A and similarly $\phi \circ \psi = n_{A'}$ on A'. *[Hint:* $\ker\phi \subseteq A[n]$.]
(b) Suppose that A is (not necessarily principally) polarized and let $\lambda\colon A \to A^\vee$ be the map induced by (43.4.20). Using (a), show that (43.4.22) gives a well-defined involution on $\operatorname{End}(A) \otimes \mathbb{Q}$, still called the **Rosati involution** (attached to the polarized abelian variety A).

▹6. Let V be a finite-dimensional vector space over $\mathbb{C}$.

(a) Let $H\colon V \times V \to \mathbb{C}$ be a nondegenerate Hermitian form on V. Let $S := \operatorname{Re} H$ and $E := \operatorname{Im} H$, so that $H = S + iE$ for $\mathbb{R}$-bilinear forms $S, E\colon V \times V \to \mathbb{R}$. Show that S is symmetric (i.e., $S(y,x) = S(x,y)$), E is alternating (i.e., $E(y,x) = -E(x,y)$), and $S(x,y) = E(ix,y)$ for all $x, y \in V$, and moreover that E (and S) are nondegenerate.
(b) Let E be a nondegenerate alternating form on V (as an $\mathbb{R}$-vector space). Show there exists a unique nondegenerate Hermitian form H on V with $\operatorname{Im} H = E$ if and only if $E(ix, iy) = E(x, y)$ for all $x, y \in V$.
(c) Show that a Hermitian form on V is positive definite if and only if the corresponding symmetric form $S := \operatorname{Re} H$ is positive definite.

▹7. Prove Proposition 43.2.1.

▹8. In the following exercise, we do a few manipulations with generating functions, applied to understand the presentation of the ring of modular forms in the next exercise.

(a) Prove that

$$\sum_{k \in 2\mathbb{Z}_{\geq 2}} \left\lfloor \frac{k}{4} \right\rfloor t^k = \frac{t^4}{(1-t^2)^2(1+t^2)}$$
$$\sum_{k \in 2\mathbb{Z}_{\geq 2}} \left\lfloor \frac{k}{3} \right\rfloor t^k = \frac{t^4+t^6}{1-t^2-t^6+t^8}$$

[Hint: break up the sum by congruence class according to the floor and then use geometric series.]

(b) Let $m_2, m_3 \in \mathbb{Z}_{\geq 0}$. For $k \in 2\mathbb{Z}_{\geq 0}$, define

$$c_k = \begin{cases} 1, & \text{if } k = 0; \\ g, & \text{if } k = 2; \\ (k-1)(g-1) + m_2\lfloor k/4 \rfloor + m_3 \lfloor k/3 \rfloor, & \text{if } k \geq 4. \end{cases}$$

Show that

$$\sum_{k \in 2\mathbb{Z}_{\geq 0}} c_k t^k = \frac{1 + gt^2 + a_4 t^4 + a_6 t^6 + a_4 t^8 + g t^{10} + t^{12}}{(1-t^4)(1-t^6)}$$

where

$$a_4 = 3g + m_2 + m_3 - 4$$
$$a_6 = 4g + m_2 + 2m_3 - 6.$$

▹9. Prove Proposition 43.2.7 as follows.

(a) Show that the functions f_4, f_6 are algebraically independent. *[Hint: reduce to the case where the relation is weighted homogeneous, and plug in z_2 to show the relation reduces to one of smaller degree.]*

(b) Show that any relation between f_4, g_6, h_{12} is a multiple of r. *[Hint: Without loss of generality, we may suppose that r is of the form $h_{12}^2 \in \mathbb{C}[f_4, g_6]_{24}$. Therefore, another relation expresses h_{12} as a rational function in f_4, g_6. Use (a) and unique factorization to show that this purported relation is in fact polynomial, and obtain a contradiction from the linear independence of f_4^3, g_6^2, h_{12}.]*

(c) Show that the subring of $M(\Gamma^1)$ generated by f_4, g_6, h_{12} is isomorphic to

$$M' = \mathbb{C}[f_4, g_6, h_{12}]/\langle r(f_4, g_6, h_{12}) \rangle.$$

(d) Show that

$$\sum_{k \in 2\mathbb{Z}_{\geq 0}} (\dim_{\mathbb{C}} M_k(\Gamma^1)) t^k = \frac{1+t^{12}}{(1-t^4)(1-t^6)}$$

and $\dim_{\mathbb{C}} M_k = \dim_{\mathbb{C}} M'_k$ for all k. *[Hint: use Exercise 43.8.]*

(f) Conclude that the subring of $M(\Gamma^1)$ generated by f_4, g_6, h_{12} is equal to $M(\Gamma^1)$. *[Hint: Suppose that equality does not hold, and consider the minimal degree with a new generator; by dimensions, there must be a new relation, but argue that this relation must be among* f_4, g_6, h_{12}*, contradicting (b).]*

This chapter is licensed under the terms of the Creative Commons Attribution-NonCommercial 4.0 International License (http://creativecommons.org/licenses/by-nc/4.0/), which permits any noncommercial use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license and indicate if changes were made.

The images or other third party material in this chapter are included in the chapter's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the chapter's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.

Symbol Definition List

$\Vert\ \Vert$	idelic norm, 464
$\Vert\phi\Vert$	modulus of an automorphism ϕ of a topological group, 510
$\left(\dfrac{K,b}{F}\right)$, $(K,b \mid F)$	quaternion algebra defined by $K \supseteq F$ and $b \in F^\times$, 86
$\left(\dfrac{K}{\mathfrak{p}}\right)$	splitting (Legendre) symbol, 552
$\left(\dfrac{O}{\mathfrak{p}}\right)$, $(O \mid \mathfrak{p})$	Eichler symbol of a local, 400
$(a,b)_F$	Hilbert symbol on $a,b \in F^\times$, 193
$\left(\dfrac{a,b}{F}\right)$, $(a,b \mid F)$	quaternion algebra defined by $a,b \in F^\times$ (char $F \neq 2$), 22
$\overline{}$	involution, 36
$*$, $\dagger$	positive involution, 128
⌑	fundamental set, 631
$*$	conjugate transpose (of a complex matrix), 27
$\langle a_1,\dots,a_n\rangle$	diagonal quadratic form on $a_1,\dots,a_n$, 51
$\langle\ \rangle$	additive homomorphism of a local field, 515
$\left[\dfrac{a,b}{F}\right)$, $[a,b \mid F)$	quaternion algebra defined by $a,b \in F^\times$, char $F = 2$, 86
$\vert\ \vert$	absolute value, 185
$\vert\ \vert_p$	p-adic absolute value, 181
$2D_n$	binary dihedral (dicyclic) group of order $2n$, 176
$2I$	binary icosahedral group, 175
$2O$, $\widetilde{S}_4$	binary octahedral group, 174
$2T$, $\widetilde{A}_4$	binary tetrahedral group, 168
$[M:N]_R$	index of R-lattices $M,N \subseteq V$, 146
α,β	elements (of an algebra), 23
A	F-algebra, 39
$\mathfrak{a},\mathfrak{b}$	fractional ideals (of a noetherian domain), 140
$a_{\mathfrak{n}}(O)$	count of right O-ideals of reduced norm $\mathfrak{n}$, 436
$\mathrm{AL}(O)$	the Atkin–Lehner group of the R-order O, 496
$\widehat{\alpha},\widehat{\beta}$	adelic or idelic elements of an algebra, 462
A_n	alternating group on n letters, 167

© The Editor(s) (if applicable) and The Author(s) 2021

J. Voight, *Quaternion Algebras*, Graduate Texts in Mathematics 288,
https://doi.org/10.1007/978-3-030-56694-4

$\widehat{E} = \widehat{E}_{\widehat{S},\widehat{O}}$	elements conjugating a fixed embedding into adelic optimal embeddings $\widehat{S} \hookrightarrow \widehat{O}$, 548
E_k, G_k	Eisenstein series, 745
$\mathrm{Emb}(\widehat{S}, \widehat{O}; \widehat{\Gamma})$	set of $\widehat{\Gamma}$-conjugacy classes of optimal adelic embeddings $\widehat{S} \hookrightarrow \widehat{O}$ of the quadratic order $\widehat{S}$ into the quaternion order $\widehat{O}$, 548
$\mathrm{Emb}_R(S, O)$	set of optimal embeddings $S \hookrightarrow O$ of the quadratic order S into the quaternion order O, 546
$\mathbf{E}^n$	Euclidean n-space, 606
$\mathrm{Emb}(S, O; \Gamma)$	set of Γ-conjugacy classes of optimal embeddings $S \hookrightarrow O$, 546
F	field, 21
$\underline{F}^{(1)}$	product norm 1 ideles, 464
F^{al}	algebraic closure of the field F, 21
$f^{\vee}$	Fourier transform of Schwartz function f, 504
$\mathfrak{f}(O)$	conductor of the quaternion R-order O, 397
$F^{\times}_{>_{\Omega}0}$	elements of F positive at the real places $v \in \Omega$, 233
F^{sep}	separable closure of F, 93
$\mathrm{Gal}(K \mid F)$	Galois group of the Galois extension of fields $K \supseteq F$, 24
$\Gamma_0(N), \Gamma_1(N)$	standard congruence subgroups of level N, 658
$\Gamma(N)$	full congruence subgroup of level N, 658
$\Gamma_{\mathbb{R}}(s), \Gamma_{\mathbb{C}}(s)$	standard archimedean Γ-factors, 448
$\mathrm{Gen}\, O$	genus of an R-order O, 282
$\mathrm{Gen}\, Q$	genus of a quadratic module Q, 149
$GN(O)$	idelic reduced norm group attached to the normalizer of O, 485
$G(O)$	idelic reduced norm group attached to O, 473
$\mathrm{GO}(Q)(F)$	general orthogonal group of Q, 49
$\mathrm{Gor}(O)$	Gorenstein saturation of the quaternion R-order O, 397
G	linear algebraic group, 723
$\mathrm{GSO}(Q)(F)$	general special orthogonal group of a quadratic form Q, 55
$\mathbb{H}$	(real) Hamiltonians, 23
$\mathbb{H}^0$	pure Hamiltonians, 28
$\mathbb{H}^1$	unit Hamiltonians, 27
$\mathbf{H}^2$	upper half-plane, 608
$\mathbf{H}^{2\pm}$	union of upper and lower half-planes, 609
$\mathbf{H}^{2*}$	completed upper half-plane, 610
$\mathbf{H}^3$	upper half-space, 663
$\mathcal{H}$	product of hyperbolic planes and spaces, 718
$\mathbf{H}^n$	hyperbolic n-space, 670
H	hyperbolic plane, 73
$h(S)$	class number of S, 548
I	lattice (over a domain, in an algebra), 156
I_2, I_4, I_6, I_{10}	Igusa invariants, 816
$I(\alpha)$	axis of $\alpha \in \mathbb{H}^1$, 29

$\mathrm{Idl}(O)$	group of invertible two-sided fractional O-ideals, 298
I^{-1}	quasi-inverse of a lattice I, 265
i, j	standard generators for a quaternion algebra, 23
int	interior, 631
$I^\sharp$	dual of the R-lattice I with respect to the reduced trace, 250
$\jmath$	automorphy factor, 749
K	$K \supseteq F$ quadratic algebra, 39
k	residue field, 186
L	projective R-module of rank 1, 149
L^2	Lorentz hyperboloid, 620
L^3	Lorentz model of hyperbolic 3-space, 670
Λ	lattice, 657
λ	left regular representation, 24
$\lambda(O, \mathfrak{p})$	mass correction term at a prime $\mathfrak{p}$ for the quaternion order O, 423
$L^\vee$	dual of the the invertible R-module L, 245
$\mathcal{L}$	Lobachevsky function, 672
$\ell(\upsilon)$	length of the path upsilon, 606
$\mathfrak{m}$	maximal ideal (of a domain), 143
M, N	lattices (over a domain, inside a vector space), 141
$m(\alpha_1, \alpha_2, \alpha_3)$	box product of elements in a quaternion algebra, 247
$\mathrm{mass}(\mathrm{Cls}\, O)$	mass of the class set of a quaternion order O, 432
$\mathrm{M}_n(A)$	ring of $n \times n$-matrices over a ring A, 21
$M_{(\mathfrak{p})}$	localization of an R-lattice M at $\mathfrak{p}$, 143
$M_\mathfrak{p}$	completion of an R-lattice M at $\mathfrak{p}$, 146
$m(\widehat{S}, \widehat{O}; \widehat{\Gamma})$	number of $\widehat{\Gamma}$-conjugacy classes of optimal adelic embeddings $\widehat{S} \hookrightarrow \widehat{O}$, 548
$m(S, O; \Gamma)$	number of Γ-conjugacy classes of optimal embeddings $S \hookrightarrow O$, 547
$\underline{\mu}$	adelic (product) measure, 521
μ	polarization (on a quaternion order), 817
$\underline{\mu}^\times$	idelic (product) measure on $\underline{\mathbb{Q}}^\times$, 506
μ_p	standard Haar measure on $\mathbb{Q}_p$, 505
$\mu_p^\times$	normalized Haar measure on $\mathbb{Q}_p^\times$, 506
N	absolute norm, 264
$N_{B^\times}$	normalizer in $B^\times$, 177
Nm	algebra norm, 39
nrd	reduced norm, 38
O	order (in an algebra), 155
$\widehat{O}$	profinite completion of the order O over a global ring, 462
$\widehat{O}^\times$	finite idele group of the order O over a global ring, 462
$O_\mathsf{L}(I), O_\mathsf{R}(I)$	left, right order of a lattice I (in an algebra), 155
Ω	set of real, ramified places of a quaternion algebra B, 233

$\mathrm{O}(n)$	orthogonal group, 29
$\mathcal{O}^{\natural}$	radical idealizer of the local R-order $\mathcal{O}$, 403
$\mathrm{O}(Q)(F)$	orthogonal group of Q, 49
$\mathfrak{p}$	prime ideal (of a domain), 143
P	maximal (two-sided) ideal of valuation ring, 201
$\mathbb{P}^1(A)$	projective line over a commutative ring A, 439
$\mathrm{P}G$	the group G modulo scalars, 594
φ	Euler totient function, 415
Π	big period matrix, 809
π	uniformizer, 186
Pic	Picard group, 298
$\mathrm{Pic}(\mathcal{O},\mathcal{O}')$	set of homothety classes of lattices with left, right order $\mathcal{O},\mathcal{O}'$, 313
$\mathrm{Pic}_R(\mathcal{O})$	Picard group of the R-order $\mathcal{O}$, 303
$\mathrm{PIdl}(\mathcal{O})$	group of principal two-sided fractional $\mathcal{O}$-ideals, 298
ϖ	nontrivial normalizer of an Eichler order, 376
$\mathrm{Pl}\,F$	set of places of a global field F, 227
$\mathfrak{p}$	maximal ideal (of a valuation ring), 186
$\wp$	Artin–Schreier group, 191
Ψ	standard function on the algebra B, the characteristic function of a maximal order, 518
ψ	Dedekind ψ-function, 423
ψ	standard unitary character, 512
$\underline{\mathbb{Q}}^{\times}$	rational idele group, 460
$\mathbb{Q}$	field of rational numbers, 21
$\mathbb{Q}_S$	product of completions of $\mathbb{Q}$ at S, 459
Q	quadratic form, 49
Q_8	quaternion group of order 8, 166
$\widehat{\mathbb{Q}}$	profinite completion of $\mathbb{Q}$, 460
$\underline{\mathbb{Q}}_{\not S}$	projection of rational adele ring away from S, 459
$\mathbb{Q}_p$	field of p-adic numbers, 181
$\underline{\mathbb{Q}}$	rational adele ring, 458
$\mathbb{R}$	field of real numbers, 21
R	(commutative) noetherian domain, 140
R	valuation ring, 186
rad	radical of a quadratic form, 53
$\mathrm{rad}\,A$	Jacobson radical of a ring A, 326
$\mathrm{rad}\,B$	Jacobson radical of an algebra B, 103
$\mathrm{Ram}\,B$	ramification set of a quaternion algebra B, 218
rev	reversal map (on a Clifford algebra), 71
ρ	distance (in a metric space), 606
$R_{(\mathfrak{p})}$	localization of a domain R at $\mathfrak{p}$, 143
$R_{\mathfrak{p}}$	completion of a domain R at $\mathfrak{p}$, 146
$R_{(S)}$	ring of S-integers of a global field, 229

S	set of places of a global field, 229
S	quadratic R-algebra, 206
$\operatorname{sgndisc}(Q)$	signed discriminant of a quadratic form Q, 52
S_n	symmetric group on n letters, 166
$\mathrm{SO}(n)$	special orthogonal group, 29
$\mathrm{SO}(Q)(F)$	special orthogonal group of the quadratic form Q, 55
$\mathrm{Stab}_G(x)$	stabilizer of $x \in X$ under the action by the group G, 630
$\mathrm{StCl}\, O$	stable (or locally free) class group, 337
$\mathrm{SU}(n)$	special unitary group, 27
$\mathcal{T}$	Bruhat–Tits tree, 388
T	associated (symmetric) bilinear form, 49
$\underline{\tau}$	Tamagawa measure, 522
τ	self-dual measure, 511
τ_x	reflection in x, 56
$\mathrm{Ten}^0\, V$	even tensor algebra of a vector space V, 71
$\mathrm{Ten}\, V$	tensor algebra of a vector space V, 69
Θ	Jacobi theta function, 504
$\vartheta(⌑, v), \vartheta(⌑, \ell)$	interior angle of ⌑ at the vertex v or edge ℓ, 700
$T(\mathfrak{n})$	$\mathfrak{n}$-Brandt matrix, 763
Tr	algebra trace, 39
trd	reduced trace, 38
$\mathrm{Typ}\, O$	type set of an R-order O, 282
$^{\mathsf{t}}$	transpose (of a matrix), 27
υ	path, 606
U_S	open subset of the adeles, 458
V	finite-dimensional vector space, 141
v, w	valuation, 186
$⌑(\Gamma; z_0)$	Dirichlet domain for Γ centered at z_0, 685
w_J	unit group modular scalars of the left order of the R-lattice J, 415
ω	Hurwitz element, 165
$\wp$	Weierstrass $\wp$-function, 747
X	metric space, 606
$\xi_B(s)$	completed zeta function of the quaternion algebra B, 449
$\mathbb{Z}$	ring of integers, 21
$Z(B)$	center of an algebra B, 21
$Z_B^{\Phi}(s)$	local zeta function of the algebra B with respect to the function Φ, 518
ζ	orientation of a quadratic form, 78
$\zeta_B(s)$	zeta function of the quaternion algebra B, 421
$\zeta_{K,[\mathfrak{b}]}(s)$	partial zeta function of the number field K, 417
$\zeta_K(s)$	Dedekind zeta function of the number field K, 417
$\zeta_O(s)$	zeta function of the quaternion order O, 420
$\zeta(s)$	Riemann zeta function, 416

Bibliography

[AB2008] Abramenko, P., & Brown, K. S. (2008). *Buildings: Theory and applications*. Graduate Texts in Mathematics (Vol. 248). New York: Springer. 23.5.17

[AA66] Adams, J. F., & Atiyah, M. F. (1966). K-theory and the Hopf invariant. *Quarterly Journal of Mathematics Oxford (2)*, *17*, 31–38. 1.2

[Alb34] Albert, A. A. (1934). Integral domains of rational generalized quaternion algebras. *Bulletin of the American Mathematical Society*, *40*, 164–176. 1.3, 14.2.8

[Alb39] Albert, A. A. (1939). *Structure of algebras*, AMS Colloquium Publications (Vol. 24). Providence, RI: American Mathematical Society. 1.3, 8.2.9

[Alb72] Albert, A. A. (1972). Tensor products of quaternion algebras. *Proceedings of the American Mathematical Society*, *35*(1), 65–66. 8.2.9

[AH32] Albert, A. A., & Hasse, H. (1932). A determination of all normal division algebras over an algebraic number field. *Transactions of the American Mathematical Society*, *34*(3), 722–726.

[Alp93] Alperin, R. C. (1993). $PSL_2(\mathbb{Z}) = \mathbb{Z}_2 * \mathbb{Z}_3$. *American Mathematical Monthly*, *100*(4), 385–386. 35.1.20

[AB2004] Alsina, M., & Bayer, P. (2004). *Quaternion orders, quadratic forms, and Shimura curves*, CRM Monograph Series (Vol. 22). American Mathematical Society: Providence, RI. 1.4, 14.2.13, 37.9.13

[Alt89] Altmann, S. L. (1989). Hamilton, Rodrigues, and the quaternion scandal. *Mathematics Magazine*, *62*(5), 291–308. 1.1

[AO2005] Altmann, S. L., & Ortiz, E. L. (Eds.). (2005). *Mathematics and social utopias in France: Olinde Rodrigues and his times*. Providence, RI: American Mathematical Society. 1.1

[ART79] Amitsur, S. A., Rowen, L. H., & Tignol, J. P. (1979). Division algebras of degree 4 and 8 with involution. *Israel Journal of Mathematics*, *33*, 133–148. 8.3.6

[Apo76] Apostol, T. M. (1976). *Introduction to analytic number theory*, Undergraduate Texts in Mathematics. New York-Heidelberg: Springer. 14.2

[Apo90] Apostol, T. M. (1990). *Modular functions and Dirichlet series in number theory*. New York: Springer. 35, 40.1

© The Editor(s) (if applicable) and The Author(s) 2021

J. Voight, *Quaternion Algebras*, Graduate Texts in Mathematics 288,
https://doi.org/10.1007/978-3-030-56694-4

[And2005] Anderson, J. W. (2005). *Hyperbolic geometry* (2nd ed.). London: Springer. 33.1

[A-C2012] Arenas-Carmona, L. (2012). Maximal selectivity for orders in fields. *Journal of Number Theory*, *132*(12), 2748–2755. 31.7.7

[A-C2013] Arenas-Carmona, L. (2013). Eichler orders, trees and representation fields. *International Journal of Number Theory*, *9*(7), 1725–1741. 31.7.7

[Art26] Artin, E. (1926). Zur Theorie der hyperkomplexen Zahlen. *Abhandlungen aus dem Mathematischen Seminar der Hamburgischen Universität*, *5*, 251–260. 7.3.11

[AT2008] Arhangel'skii, A., & Tkachenko, M. (2008). *Topological groups and related structures*, Atlantist Studies in Mathematics (Vol. 1). Hackensack, NJ: World Scientific Publishing Co. 34.1

[Arm88] Armstrong, M. A. (1988). *Groups and symmetry*, Undergraduate Texts in Mathematics. New York: Springer. 11.5

[Art50] Artin, E. (1950). The influence of J. H. M. Wedderburn on the development of modern algebra. *Bulletin of the American Mathematical Society*, *56*(1), 65–72. 1.2

[Asl96] Aslaksen, H. (1996, Summer). Quaternionic determinants. *The Mathematical Intelligencer*, *18*(3), 57–65. 2.20

[AM69] Atiyah, M. F., & Macdonald, I. G. (1969). *Introduction to commutative algebra*. Reading: Addison-Wesley. 9.2

[Ati67] Atiyah, M. F. (1967). *K-theory*. New York: W. A. Benjamin Inc.

[ABB2014] Auel, A., Bernardara, M., & Bolognesi, M. (2014). Fibrations in complete intersections of quadrics, Clifford algebras, derived categories, and rationality problems. *Journal de Mathématiques Pures et Appliquées (9)*, *102*(1), 249–291. 22.2.14

[ABGV2006] Auel, A., Brussel, E., Garibaldi, S., & Vishne, U. (2011). Open problems on central simple algebras. *Transformation Groups*, *16*(1), 219–264. 7.5.5

[Auel2011] Auel, A. (2011, January). Clifford invariants of line bundle-valued quadratic forms. *MPIM preprint series, no. 33*. 22.2.14, 22.2.15

[Auel2015] Auel, A. (2015). Surjectivity of the total Clifford invariant and Brauer dimension. *Journal of Algebra*, *443*, 395–421.

[AG60] Auslander, M., & Goldman, O. (1960). Maximal orders. *Transactions of the American Mathematical Society*, *97*, 1–24. 15.17, 18.3

[BG2008] Baba, S., & Granath, H. (2008). Genus 2 curves with quaternionic multiplication. *Canadian Journal of Mathematics*, *60*(4), 734–757. 43.2.10, 43.2.13, 43.2.15, 43.2.16, 43.2.20, 43.2.21

[Bae2002] Baez, J. C. (2002). The octonions. *Bulletin of the American Mathematical Society (N.S.)*, *39*, 145–205; Errata for "The octonions". *Bulletin of the American Mathematical Society (N.S.)*, *42*, 213 (2005). 1.2

[Bae2005] Baez, J. C. (2005). On quaternions and octonions: Their geometry, arithmetic, and symmetries by John H. Conway and Derek A. Smith. *Bulletin of the American Mathematical Society*, *42*, 229–243. 11.1

[Bak71] Baker, A. (1966). Linear forms in the logarithms of algebraic numbers. *Mathematika*, *13*, 204–216. 25.2.19

[Bal2007] Balaji Thiruvalloor Eesanaipaadi, V. (2007). Line-bundle-valued ternary quadratic forms over schemes. *Journal of Pure and Applied Algebra*, *208*, 237–259. 22.5.13

[Bas62] Bass, H. Injective dimension in Noetherian rings. *Transactions of the American Mathematical Society*, *102*. 24.2.24

[Bas63] Bass, H. (1963). On the ubiquity of Gorenstein rings. *Mathematische Zeitschrift*, *82*, 8–28. 24.2.24

[BHV2008] Bekka, B., de la Harpe, P., & Valette, A. (2008). *Kazhdan's property (T)*. Cambridge: Cambridge University Press. 29.3

[BCK93] Beukers, F., Calabi, E., & Kolk, J. (1993). Sums of generalized harmonic series and volumes. *Nieuw Archief voor Wiskunde*, *11*, 217–224. 25.1

[Bea95] Beardon, A. (1995). The geometry of discrete groups. *Graduate texts in mathematics* (Vol. 91). New York: Springer. 33.1, 34.7, 37.3

[BP92] Benedetti, R., & Petronio, C. (1992). *Lectures on hyperbolic geometry*, Universitext. Berlin: Springer. 36.3.17

[BO2013] Berhuy, G., Oggier, F. (2013). An introduction to central simple algebras and their applications to wireless communication. *Mathematical surveys and monographs* (Vol. 191). Providence, RI: American Mathematical Society. 1.5

[BK2000] Berrick, A. J., & Keating, M. E. (2000). An introduction to rings and modules with K-theory in view. *Cambridge studies in advanced mathematics* (Vol. 65). Cambridge: Cambridge University Press. 20.2

[Bha2004a] Bhargava, M. (2004). Higher composition laws. I. A new view on Gauss composition, and quadratic generalizations. *Annals of Mathematics (2)*, *159*, 217–250. 19.1

[Bia1892] Bianchi, L. (1892). Sui gruppi di sostituzioni lineari con coefficienti appartenenti a corpi quadratici immaginari. *Mathematische Annalen*, *40*, 332–412. 36.4.4

[Bic85] Bichsel, W. (1985). *Quadratische raeume mit werten in invertierbaren moduln*, Ph.D. thesis, ETH Zürich. 22.2.14

[BK94] Bichsel, W., & Knus, M.-A. (1994). *Quadratic forms with values in line bundles*, Recent advances in real algebraic geometry and quadratic forms (Berkeley, CA, 1990, 1991; San Francisco, CA, 1991). Contemporary Mathematics (Vol. 155, pp. 293–306). Providence, RI: American Mathematical Society. 22.2.13, 22.2.14

[BL2004] Birkenhake, C., & Lange, H. (2004). *Complex abelian varities* (Second augmented ed.). Berlin: Springer. 8.5, 43.4

[Bol1887] Bolza, O. (1887). On binary sextics with linear transformations into themselves. *American Journal of Mathematics*, *10*(1), 47–70. 43.3.4

[Bor62] Borel, A. (1962). Arithmetic properties of linear algebraic groups. In *Proceedings of the International Congress of Mathematicians* (Stockholm, 1962), Institute Mittag-Leffler, Djursholm, 1963 (pp. 10–22). 38.5.1

[Bor69] Borel, A. (1969). *Introduction aux groupes arithmétiques*, Publications de l'Institut de Mathématique de l'Université de Strasbourg, vol. XV, Actualités Scientifiques et Industrielles, No. 1341, Hermann, Paris. 38.5.1

[Bor81] Borel, A. (1981). Commensurability classes and volumes of hyperbolic 3-manifolds. *The Annali della Scuola Normale Superiore di Pisa, Classe di Scienze (4)*, *8*(1), 1–33.

[BHC62] Borel, A. (1962). Harish-Chandra, Arithmetic subgroups of algebraic groups. *Annals of Mathematics (2)*, *75*, 485–535. 38.5.6

[BS66] Borevich, Z. I., & Shafarevich, I. R. (1966). *Number theory*. New York: Academic Press. 26.2, 26.2.13, 26.2.19

[BM58] Bott, R., & Milnor, J. (1958). On the parallelizability of the spheres. *Bulletin of the American Mathematical Society*, *64*, 87–89. 1.2

[Bou60] Bourbaki, N. (1966). *Elements of mathematics: General topology*, Part 1. Reading: Addison-Wesley. 34.5.4

[Bou98] Bourbaki, N. (1998). *Commutative algebra, Chapters 1–7*. Berlin: Springer. 9.2

[Bra13] Brandt, H. (1913). Zur Komposition der quaternären quadratischen Formen. *Journal für die reine und angewandte Mathematik*, *143*, 106–129. 19.1

[Bra24] Brandt, H. (1924). Der Kompositionsbegriff bei den quaternären quadratischen Formen. *Mathematische Annalen*, *91*(3–4), 300–315. 19.1

[Bra25] Brandt, H. (1925). Die Hauptklassen in der Kompositionstheorie der quaternären quadratischen Formen. *Mathematische Annalen*, *94*(1), 166–175. 19.1

[Bra27] Brandt, H. (1927). Über eine Verallgemeinerung des Gruppenbegriffes. *Mathematische Annalen*, *96*, 360–366. 19.3.11

[Bra28] Brandt, H. (1928). Idealtheorie in Quaternionenalgebren. *Mathematische Annalen*, *99*, 1–29. 1.4, 17.1, 19.1, 25.1

[Bra37] Brandt, H. (1937). Zur Zahlentheorie der quadratische Formen. *Jahresbericht der Deutschen Mathematiker-Vereinigung*, *47*, 149–159. 19.1

[Bra40] Brandt, H. (1940). Uber die Axiome des Gruppoids. *Vierteljahrsschrift der Naturforschenden Gesellschaft in Zürich*, *85*, 95–104. Beiblatt (Festschrift Rudolph Fueter). 1.4, 19.3.11

[Bra41] Brandt, H. (1941). Uber die Zerlegungsgesetze der rationalen Zahlen in Quaternionen-Körpern. *Mathematische Annalen*, *117*, 758–763. 1.4

[Bra43] Brandt, H. (1943). Zur Zahlentheorie der Quaternionen. *Jahresbericht der Deutschen Mathematiker-Vereinigung*, *53*, 23–57. 1.4, 19.1, 22.5.13, 41.5.11

[BHN31] Brauer, R., Hasse, H., & Noether, E. (1932). Beweis eines Hauptsatzes in der Theorie der Algebren. *Journal für die reine und angewandte Mathematik*, *167*, 399–404.

[Bre2010] Brešar, M. (2010). An elementary approach to Wedderburn's structure theory. *Expositiones Mathematicae*, *28*, 79–83. 7.2

[BH99] Bridson, M. R., & Haefliger, A. (1999). *Metric spaces of non-positive curvature*, Grundlehren der Mathematischen Wissenschaften (Vol. 319). Berlin: Springer. 33.2

[Bro61] Brown, B. H. (1961, November 3). *Mathematics at Dartmouth: 1769–1961*, Dedicatory Conference, Albert Bradley Center for Mathematics, Dartmouth College, Hanover.

[Bro87] Brown, R. (1987). From groups to groupoids: A brief survey. *The Bulletin of the London Mathematical Society*, *19*, 113–134. 19.3.11

[Bruc71] Bruck, R. H. (1971). *A survey of binary systems, third printing*. Berlin: Springer. 19.3.11

[Brum63a] Brumer, A. (1963). *The structure of hereditary orders*, Ph.D. thesis, Princeton. 21.4

[Brum63b] Brumer, A. (1963). Structure of hereditary orders. *Bulletin of the American Mathematical Society*, *69*, 721–724; Addendum to "Structure of hereditary orders". *Bulletin of the American Mathematical Society*, *70*, 185 (1964). 21.4

[Brz80] Brzezinski, J. (1980). Arithmetical quadratic surfaces of genus 0, I. *Mathematica Scandinavica*, *46*, 183–208. 22.5.13, 24.3.11

[Brz82] Brzezinski, J. (1982). A characterization of Gorenstein orders in quaternion algebras. *Mathematica Scandinavica*, *50*(1), 19–24. 15.11, 16.8, 22.5.13, 24.2.24

[Brz83a] Brzezinski, J. (1983). On orders in quaternion algebras. *Communications in Algebra*, *11*(5), 501–522. 23.5, 24.1.5, 24.2.18, 24.5, 24.9

[Brz83b] Brzezinski, J. (1983). Spinor class groups of orders. *Journal of Algebra*, *84*, 468–481. 20.7.23, 22.5.13

[Brz85] Brzezinski, J. (1985). *Brauer-Severi schemes of orders*, Orders and their applications (Oberwolfach, 1984). Lecture Notes in Mathematics (Vol. 1142, pp. 18–49). Berlin: Springer. 22.5.13, 24.3.11

[Brz87] Brzezinski, J. (1987). Riemann-Roch theorem for locally principal orders. *Mathematische Annalen*, *276*, 529–536. 24.2.24

[Brz89] Brzezinski, J. (1989). A combinatorial class number formula. *Journal für die reine und angewandte Mathematik*, *402*, 199–210. 30.4.22

[Brz90] Brzezinski, J. (1990). On automorphisms of quaternion orders. *Journal für die reine und angewandte Mathematik*, *403*, 166–186. 24.1.5, 24.5.7, 24.5, 25.3, 26.1.14, 30.6.18

[Brz91] Brzezinski, J. (1991). On embedding numbers into quaternion orders. *Commentarii Mathematici Helvetici*, *66*, 302–318. 30.6.18

[Brz95] Brzezinski, J. (1995). Definite quaternion orders of class number one. *Journal de Théorie des Nombres de Bordeaux*, *7*, 93–96. 25.4

[Brz97] Brzezinski, J. (1997). A generalization of Eichler's trace formula. *Collectanea Mathematica*, *48*, 53–61. 30.9.12

[Brz98] Brzezinski, J. (1998). On traces of the Brandt-Eichler matrices. *Journal de Théorie des Nombres de Bordeaux*, *10*(2), 273–285. 25.4

[B+2003] Bump, D., Cogdell, J. W., de Shalit, E., Gaitsgory, D., Kowalski, E., & Kudla, S. S. (2003). *An introduction to the Langlands program*, eds. J. Bernstein and S. Gelbart. Boston: Birkhäuser. 1.4

[BBI2001] Burago, D., Burago, Y., & Ivanov, S. (2001). *A course in metric geometry*, Graduate Studies in Mathematics (Vol. 33). American Mathematical Society. 33.2.8

[Bus55] Busemann, H. (1955). *The geometry of geodesics*. New York: Academic Press Inc. 33.2.8

[BFS2008] Buser, P., Flach, N., & Semmler, K.-D. (2008). Spectral rigidity and discreteness of 2233-groups, with an appendix by Colin MacLachlan and Gerhard Rosenberger. *Mathematical Proceedings of the Cambridge Philosophical Society*, *144*(1), 145–180. 31.8

[BR85] Bushnell, C. J., Reiner, I. (1985). *A survey of analytic methods in noncommutative number theory*, Orders and their applications (Oberwolfach, 1984). Lecture Notes in Mathematics (Vol. 1142, pp. 50–87). Berlin: Springer. 26.3.14

[CFKP97] Cannon, J. W., Floyd, W. J., Kenyon, R., & Parry, W. R. (1997). *Hyperbolic geometry*, Flavors of geometry, ed. Silvio Levy, MSRI Publications (Vol. 31, pp. 59–115). Cambridge: Cambridge University Press. 33.1, 36.3.17

[CNP2005] Cardona, G., Nart, E., & Pujolas, J. (2005). Curves of genus two over fields of even characteristic. *Mathematische Zeitschrift*, *250*, 177–201. 43.3.10

[CQ2005] Cardona, G., & Quer, J. (2005). *Field of moduli and field of definition for curves of genus 2*, Computational aspects of algebraic curves, Lecture Notes Series Computer (Vol. 13, pp. 71–83). Hackensack, NJ: World Scientific Publishing. 43.3.10, 43.3.13

[Cas78] Cassels, J. W. S. (1978). *Rational quadratic forms*. London: Academic Press. 4.2, 14.4.4, 14.4.12

[Cas86] Cassels, J. W. S. (1986). *Local fields*, London Mathematical Society Student Texts (Vol. 3). Cambridge: Cambridge University Press. 12.2, 13.2

[Cas97] Cassels, J. W. S. (1997). *An introduction to the geometry of numbers*. Berlin: Springer. 17.5

[Cas2010] Cassels, J. W. S. (2010). *Global fields*, Algebraic number theory (2nd ed., pp. 42–84), eds. J. W. S. Cassels and A. Frölich. London: London Mathematical Society. 14.4, 27.4, 27.4.16

[Cay1845a] Cayley, A. (1845). On certain results relating to quaternions. *Philosophical Magazine*, *26*, 141–145. 1.2

[Cay1845b] Cayley, A. (1845). On Jacobi's elliptic functions, in reply to the Rev. B. Bronwin; and on quaternions. *Philosophical Magazine*, *26*, 208–211. 1.2

[CCL2013] Cerri, J.-P., Chaubert, J., & Lezowski, P. (2013). Euclidean totally definite quaternion fields over the rational field and over quadratic number fields. *International Journal of Number Theory*, *9*(3), 653–673. 26.7.4

[CX2004] Chan, W. K., & Xu, F. (2004). On representations of spinor genera. *Compositio Mathematica*, *140*(2), 287–300. 31.7, 31.7.7

[CDL2015] Chapman, A., Dolphin, A., & Laghribi, A. (2016). Total linkage of quaternion algebras and Pfister forms in characteristic two. *Journal of Pure and Applied Algebra*, *220*, 3676–3691. 8.2

[Cha2020] Chari, S. (2020). Metacommutation of primes in central simple algebras. *Journal of Number Theory*, *206*, 296–309. 11.4.11

[CGL2009] Charles, D., Goren, E., & Lauter, K. (2009). Families of Ramanujan graphs and quaternion algebras, Groups and Symmetries. *CRM Proceedings Lecture Notes*, *47*, 53–80. 41.2.14

[Chev34] Chevalley, C. (1934). Sur certains idéaux dans un algèbre simple. *Abhandlungen aus dem Mathematischen Seminar der Universität Hamburg*, *10*, 83–105. 30.3.17

[Chev36] Chevalley, C. (1936). L'arithmétique dans les algèbres de matrices. *Actualités Sci. Ind.*, *323*.

[CSV2019] Chari, S., Smertnig, D., & Voight, J. (2019). *On basic and Bass quaternion orders*. arXiv:1903.00560. 24.5

[Chi2009] Childress, N. (2009). *Class field theory*, Universitext. New York: Springer. 27

[CF99] Chinburg, T., & Friedman, E. (1999). An embedding theorem for quaternion algebras. *Journal of the London Mathematical Society (2)*, *60*(1), 33–44. 31.1.8, 31.7, 31.7.7

[CF2000] Chinburg, T., & Friedman, E. (2000). The finite subgroups of maximal arithmetic Kleinian groups. *Annales de l'Institut Fourier (Grenoble)*, *50*(6), 1765–1798. 32.5, 32.6, 32.6.5

[Cli1878] Clifford, W. K. (1878). Applications of Grassmann's extensive algebra. *American Journal of Mathematics*, *1*, 350–358. 1.2

[Coh93] Cohen, H. (1993). *A course in computational algebraic number theory*. Graduate Texts in Mathematics (Vol. 138). Berlin: Springer-Verlag.

[CK2015] Cohn, H., & Kumar, A. (2015). Metacommutation of Hurwitz primes. *Proceedings of the American Mathematical Society*, *143*(4), 1459–1469. 11.4.11

[Con] Conrad, K. (n.d) *A multivariable Hensel's lemma*, preprint, http://kconrad.math.uconn.edu/blurbs/gradnumthy/multivarhensel.pdf.

[CG93] Conrey, J. B., & Ghosh, A. (1993). On the Selberg class of Dirichlet series: Small degrees. *Duke Mathematical Journal*, *72*(3), 673–693. 26.3.15

[Con2012] Conway, J. B. (2012). *A course in abstract analysis*, Graduate Studies in Mathematics (Vol. 141). Providence, RI: American Mathematical Society. 3.2.11

[CSl88] Conway, J. H., & Sloane, N. J. A. (1988). *Sphere packings, lattices and groups*, Grundlehren der Mathematischen Wissenschaften (Vol. 290). New York: Springer. 11.5.7

[CSm2003] Conway, J. H., & Smith, D. A. (2003). *On quaternions and octonions: Their geometry, arithmetic, and symmetry*. Natick, MA: A. K. Peters Ltd. 1.2, 11.1, 11.2.4, 11.4.11, 11.5

[Cox89] Cox, D. A. (1989). *Primes of the form* $x^2 + ny^2$*: Fermat, class field theory, and complex multiplication*. New York: Wiley. 16.1, 16.5.19, 17.1, 30.9.9

[Coxtr40] Coxeter, H. S. M. (1940). The binary polyhedral groups and other generalizations of the quaternion group. *Duke Mathematical Journal*, *7*, 367–379. 32.4

[Cro64] Crowe, M. J. (1964). *A history of vector analysis*. Mineola, NY: Dover Press. 1.1, 1.2

[CR62] Curtis, C. W., & Reiner, I. (1962). *Representation theory of finite groups and associative algebras*. New York: Wiley. 9.3

[CR81] Curtis, C. W., & Reiner, I. (1981). *Methods of representation theory, Volume I: With applications to finite groups and orders*. New York: Wiley. 7.2, 7.2.20, 7.9, 9.2, 10.2, 17.3.7, 17.7.27, 18.3.9, 20.1.4, 20.2, 20.3.6, 20.4, 20.6, 22.2, 24.4, 24.5

[CR87] Curtis, C. W., & Reiner, I. (1987). *Methods of representation theory, Volume II: With applications to finite groups and orders*. New York: Wiley. 12.4.4, 18.4.11, 20.1.4, 20.7, 20.7.13

[DT2008] Dasgupta, S., & Teitelbaum, J. (2008). *The p-adic upper half plane*, *p*-adic geometry, University Lecture Series (Vol. 45, pp. 65–121). Providence, RI: American Mathematical Society. 23.5

[Dei2005] Deitmar, A. (2005). *A first course in harmonic analysis*, Universitext (2nd ed.). New York: Springer. 29.3

[DE2009] Deitmar, A., & Echterhoff, S. (2009). *Principles of harmonic analysis*, Universitext. New York: Springer. 29.3, 29.3.3, 29.3.10, 29.4

[Del71] Deligne, P. (1971). *Travaux de Shimura*, Séminaire Bourbaki, 23ème année (1970/71), Exp. No. 389, 123–165, Lecture Notes in Mathematics (Vol. 244). Berlin: Springer. 38.7.20, 43.8.7

[Del74] Deligne, P. (1974). La conjecture de Weil. I. *Publications Mathématiques de l'IHÉS*, *43*, 273–307. 41.2.14

[DR80] Deligne, P., & Ribet, K. A. (1980). Values of abelian *L*-functions at negative integers over totally real fields. *Inventiones Mathematicae*, *59*, 227–286. 26.1.9

[Dem2007] Dembélé, L. (2007). Quaternionic Manin symbols, Brandt matrices and Hilbert modular forms. *Mathematics of Computation*, *76*(258), 1039–1057. 41.5.12

[DD2008] Dembélé, L., & Donnelly, S. (2008). *Computing Hilbert modular forms over fields with nontrivial class group*, Algorithmic number theory (Banff, 2008). Lecture Notes in Computer Science (Vol. 5011, pp. 371–386). Berlin: Springer. 41.5.12

[DV2013] Dembélé, L., & Voight, J. (2013). *Explicit methods for Hilbert modular forms*, Elliptic curves, Hilbert modular forms and Galois deformations (pp. 135–198). Basel: Birkhauser. 41.5.12

[dR71] de Rham, G. (1971). Sur les polygons générateurs de groups Fuchsiens. *Enseignement Mathématique*, *17*, 49–61. 37.6.5

[DG89] Deturck, D. M., & Gordon, C. S. (1989). Isospectral deformations II: Trace formulas, metrics, and potentials, with an appendix by Kyung Bai Lee. *Communications on Pure and Applied Mathematics*, *42*, 1067–1095. 31.8

[Deu41] Deuring, M. (1941). Die Typen der Multiplikatorenringe elliptischer Funktionenkörper. *Abhandlungen aus dem mathematischen Seminar der Hansischen Universität*, *14*, 197–272. 42.1

[Deu51] Deuring, M. (1951). Die Anzahl der Typen von Maximalordnungen einer definiten Quaternionenalgebra mit primer Grundzahl. *Über. Deutsch. Mat. Verein.*, *54*, 24–41. 30.9

[Deu68] Deuring, M. (1968). *Algebren*, Ergebnisse der Mathematik und ihrer Grenzgebiete (Vol. 41). Berlin-New York: Springer. 29.10.24

[DS2005] Diamond, F., & Shurman, J. (2005). *A first course in modular forms*, Graduate Texts in Mathematics (Vol. 228). New York: Springer. 35, 40.1, 40.3, 40.5, 40.5.12

[Dic03] Dickson, L. E. (1903). Definitions of a linear associative algebra by independent postulates. *Transactions of the American Mathematical Society*, *4*(1), 21–26. 1.2

[Dic12] Dickson, L. E. (1912). Linear algebras. *Transactions of the American Mathematical Society*, *13*(1), 59–73. 1.2

[Dic14] Dickson, L. E. (1914). Linear associative algebras and abelian equations. *Transactions of the American Mathematical Society*, *15*, 31–46. 1.2, 1.3

[Dic19] Dickson, L. E. (1919). On quaternions and their generalization and the history of the eight square theorem. *Annals of Mathematics (2)*, *20*(3), 155–171. 1.1, 1.3

[Dic23] Dickson, L. E. (1923). *Algebras and their arithmetics*. Chicago: The University of Chicago Press. 1.2, 1.3

[Dic24] Dickson, L. E. (1924). On the theory of numbers and generalized quaternions. *American Journal of Mathematics*, *46*(1), 1–16. 1.3

[Dic71] Dickson, L. E. (1971). *History of the theory of numbers, Volume II: Diophantine analysis*. New York: Chelsea. 1.4

[DS2014] Diestel, J., & Spalsbury, A. (2014). *The joys of Haar measure*, Graduate Studies in Mathematics (Vol. 150). American Mathematical Society. 29.3

[Die48] Dieudonné, J. (1948). *Sur les groupes classiques*, Actualitiés scientifique et industrielles, no. 1040, Paris, Hermann. 3.5.3

[Die53] Dieudonné, J. (1953). On the structure of unitary groups (II). *American Journal of Mathematics*, *75*, 665–678. 3.5.3

[DK95] Diek, A., Kantowski, R. (1995). *Some Clifford algebra history*, *Clifford Algebras and Spinor Structures*, Mathematics and Its Applications (Vol. 321, pp. 3–12). 1.2

[DM89] Dimitrić, R., & Goldsmith, B. (1989). The mathematical tourist: Sir William Rowan Hamilton. *The Mathematical Intelligencer*, *11*(2), 29–30. 1.1

[Dix77] Dixmier, J. (1977). *C^*-algebras, Translated from the French by Francis Jellett, North-Holland Mathematical Library* (Vol. 15). North-Holland, Amsterdam. 3.2.11

[Don2011] Donaldson, S. (2011). *Riemann surfaces*. Oxford: Oxford University Press. 33.8.6

[Dra83] Draxl, P. K. (1983). *Skew fields*. Cambridge: Cambridge University Press. 7.5

[DK68] Drozd, Yu. A., & Kirichenko, V. V. (1968). *Hereditary orders*, Math. Inst. Acad. Sci. Ukranian SSR, translated from Ukrainskii Mat. Zhurnal, **20**(2), 246–248. 21.4

[DKR67] Drozd, Yu. A., Kirichenko, V. V., & Roĭter, A. V. (1967). *On hereditary and Bass orders*, Math. USSR-Izvestija, translated from Izv. Akad. Nauk SSSR, **1**(6), 1357–1375. 21.4, 24.2.24, 24.4, 24.5

[DK94] Drozd, Y. A., & Kirichenko, V. V. (1994). *Finite dimensional algebras, with an appendix by Vlastimil Dlab, translated by Vlastimil Dlab*. Berlin: Springer. 7.2, 7.9

[EGM2008] Edixhoven, B., van der Geer, G., & Moonen, B. (2008). *Modular forms*. Modular forms on Schiermonnikoog. Cambridge: Cambridge University Press. 1.4

[Eic36] Eichler, M. (1936). Untersuchungen in der Zahlentheorie der rationalen Quaternionenalgebren. *Journal für die reine und angewandte Mathematik*, *174*, 129–159. 1.4, 23.4.22, 24.1.5, 24.5

[Eic37] Eichler, M. (1937). Bestimmung der Idealklassenzahl in gewissen normalen einfachen Algebren. *Journal für die reine und angewandte Mathematik*, *176*, 192–202. 28.6.11

[Eic38a] Eichler, M. (1938). Allgemeine Kongruenzklasseneinteilungen der Ideale einfacher Algebren über algebraischen Zahlkörper und ihre L-Reihen. *Journal für die reine und angewandte Mathematik*, *179*, 227–251. 26.8, 28.6.11

[Eic38b] Eichler, M. (1938). Über die Idealklassenzahl total definiter Quaternionenalgebren. *Mathematische Zeitschrift*, *43*(1), 102–109. 26.1.14, 30.6.18

[Eic38c] Eichler, M. (1938). Über die Idealklassenzahl hyperkomplexer Systeme. *Mathematische Zeitschrift*, *43*(1), 481–494. 28.6.11

[Eic53] Eichler, M. (1974). *Quadratische Formen und orthogonale Gruppen*, Grundlehren der mathematischen Wissenschaften (Vol. 63). New York: Springer. 1.4, 4.2, 4.2.19, 22.5.13

[Eic55-56] Eichler, M. (1955–56). *Lectures on modular correspondences*, Tata Institute of Fundamental Research, Bombay. 14.5.2, 25.3

[Eic56a] Eichler, M. (1956). Zahlentheorie, Zur, der Quaternionen-Algebren. *Journal für die reine und angewandte Mathematik*, *195*, 127–151; Berichtigung zu der Arbeit "Zur Zahlentheorie der Quaternionen-Algebren". *Journal für die reine und angewandte Mathematik*, *197*, 220 (1957). 1.4, 23.4.22, 23.1, 25.3, 26.1.14, 30.3.17, 30.6.18, 30.8.8, 30.9.12, 41.5.11

[Eic56b] Eichler, M. (1956). Über die Darstellbarkeit von Mdoulformen durch Thetareihen. *Journal für die reine und angewandte Mathematik*, *195*, 156–171. 1.4

[Eic56c] Eichler, M. (1956). Modular correspondences and their representations. *Journal of the Indian Mathematical Society (N.S.)*, *20*, 163–206. 23.4.22

[Eic58] Eichler, M. (1958). Quadratische Formen und Modulfunktionen. *Acta Arithmetica*, *4*, 217–239. 1.4

[Eic73] Eichler, M. (1973). The basis problem for modular forms and the traces of the Hecke operators, Modular functions of one variable, I (Proceedings, International Summer School, University of Antwerp, Antwerp, 1972). Lecture Notes in Mathematics (Vol. 320, pp. 75–151). Berlin: Springer; Correction to: "The basis problem for modular forms and the traces of the Hecke operators", Modular functions of one variable, IV (Proceedings, International Summer School, University of Antwerp, Antwerp, 1972), Lecture Notes in Mathematics (Vol. 476, pp. 145–147). Berlin: Springer (1975). 1.4, 23.4.22, 30.3.17, 30.6.18, 40.4, 41.5.11

[Eic77] Eichler, M. (1977). On theta functions of real algebraic number fields. *Acta Arithmetica*, *33*, 269–292. 41.5.11

[EM45] Eilenberg, S., & Mac Lane, S. (1945). General theory of natural equivalences. *Transactions of the American Mathematical Society*, *58*, 231–294. 19.3.11

[Elk98] Elkies, N. D. (1998). *Shimura curve computations*, Algorithmic number theory (Portland, OR, 1998). Lecture Notes in Computer Science (Vol. 1423, pp. 1–47). Berlin: Springer. 43.2.19

[EKM2008] Elman, R. S., Karpenko, N., & Merkurjev, A. (2008). *The algebraic and geometric theory of quadratic forms* (Vol. 56). American Mathematical Society. 6.3

[EGM98] Elstrodt, J., Grunewald, F., & Mennicke, J. (1998). *Groups acting on hyperbolic space: Harmonic analysis and number theory*. Berlin: Springer. 36.1, 40.2.24

[EP94] Epstein, D. B. A., & Petronio, C. (1994). An exposition of Poincaré's polyhedron theorem. *Enseignement Mathématique*, *40*, 113–170. 37.6, 37.6.5

[Ewe82] Ewell, J. A. (1982). A simple derivation of Jacobi's four-square formula. *Proceedings of the American Mathematical Society*, *85*(3), 323–326. 1.4

[Fad65] Faddeev, D. K. (1965). *Introduction to multiplicative theory of modules of integral representations*, Algebraic number theory and representations, Proceedings of the Steklov Institute of Mathematics (Vol. 80, pp. 164–210). Providence, RI: American Mathematical Society (1968). 15.6, 16.8, 20.3.6

[FD93] Farb, B., & Keith Dennis, R. (1993). *Noncommutative algebra*, Graduate Texts in Mathematics (Vol. 144). New York: Springer. 7.2

[Fen98] Fenster, D. D. (1998). Leonard Eugene Dickson and his work in the Arithmetics of Algebras. *Archive for History of Exact Sciences*, *52*(2), 119–159. 1.2

[FS2007] Fenster, D. D., & Schwermer, J. (2007). Beyond class field theory: Helmut Hasse's arithmetic in the theory of algebras in 1931. *Archive for History of Exact Sciences*, *61*, 425–456. 1.3, 14.6

[Fin89] Fine, B. (1989). *The algebraic theory of the Bianchi groups*. New York: Marcel Dekker. 35.5, 36.4.4, 36.8

[Fit2011] Fitzgerald, R. W. (2011). Norm Euclidean quaternionic orders. *INTEGERS*, *11*(A58). 17.10

[Fol95] Folland, G. (1995). *A course in abstract harmonic analysis*, Studies in Advanced Mathematics. Boca Raton, FL: CRC Press. 29.4

[For72] Ford, L. R. (1972). *Automorphic functions* (2nd ed.). New York: Chelsea. 33.1, 37.2.9

[FGS2016] Forsyth, A., Gurev, J., & Shrima, S. (2016). Metacommutation as a group action on the projective line over $\mathbb{F}_p$. *Proceedings of the American Mathematical Society*, *144*(11), 4583–4590.

[FK1890-2] Fricke, R., & Klein, F. (1890, 1892). *Vorlesungen über die Theorie der elliptischen Modulfuctionen* (Vols. 1 and 2). Leipzig: B.G. Teubner. 1.4, 1.4.4

[FK1897] Fricke, R., & Klein, F. (1897). *Vorlesungen über die Theorie der automorphen Functionen*. Erster Band: Die gruppentheoretischen Grundlagen. Leipzig: B.G. Teubner. 1.4

[FK12] Fricke, R., & Klein, F. (1912). *Vorlesungen über die Theorie der automorphen Functionen*. Zweiter Band: Die funktionentheoretischen Ausführungen und die Anwendungen. Leipzig: B.G. Teubner. 1.4

[Fro1878] Frobenius, F. G. (1878). Über lineare Substitutionen und bilineare Formen. *Journal für die reine und angewandte Mathematik*, *84*, 1–63. 1.2

[Frö75] Fröhlich, A. (1975). Locally free modules over arithmetic orders. *Journal für die reine und angewandte Mathematik*, *274*(275), 112–124. 20.7, 20.7.23

[Frö73] Fröhlich, A. (1973). The Picard group of noncommutative rings, in particular of orders. *Transactions of the American Mathematical Society*, *180*, 1–45. 18.4.11

[FRU74] Fröhlich, A., Reiner, I., & Ullom, S. (1974). Class groups and Picard groups of orders. *Proceedings of the London Mathematical Society (3)*, *29*, 405–434. 20.7.23

[FT91] Fröhlich, A., & Taylor, M.J. (1991). *Algebraic number theory*, Cambridge Studies in Advanced Mathematics (Vol. 27). Cambridge: Cambridge University Press. 9.3

[Fuj58] Fujisaki, G. (1958). On the zeta-function of the simple algebra over the field of rational numbers. *Journal of the Faculty of Science, University of Tokyo. Section I*, *7*, 567–604. 27.6, 29.10.24

[FH91] Fulton, W., & Harris, J. (1991). *Representation theory: A first course*, Graduate Texts in Mathematics (Vol. 129). New York: Springer. 2.17

[GV2018] Galbraith, S. D., & Vercauteren, F. (2018). Computational problems in supersingular elliptic curve isogenies. *Quantum Information Science*, *17*(10), Art. 265, 22pp. 1.5

[Gar2004] Garibaldi, S. (2004). The characteristic polynomial and determinant are not ad hoc constructions. *American Mathematical Monthly*, *111*(9), 761–778. 7.8

[GS2010] Garibaldi, S., & Saltman, D. J. (2010). *Quaternion algebras with the same subfields*, Quadratic forms, linear algebraic groups, and cohomology (pp. 225–238). New York: Springer. 8.2.6, 14.18

[Gau00] Gauss, C. F. (1900). *Mutation des Raumes*, in Carl Friedrich Gauss Werke (Vol. 8, pp. 357–361). König. Gesell. Wissen.: Göttingen. 1.1

[Gau86] Gauss, C. F. (1986). *Disquisitiones arithmeticae*, Translated and with a preface by Arthur A. Clarke, revised by William C. Waterhouse, Cornelius Greither, and A. W. Grootendorst, and with a preface by Waterhouse. New York: Springer. 25.1, 25.2.19, 30.1

[GMMR97] Gehring, F. W., Maclachlan, C., Martin, G. J., & Reid, A. W. (1997). Arithmeticity, discreteness and volume. *Transactions of the American Mathematical Society*, *349*, 3611–3643. 32.7

[Gel84] Gelbart, S. (1984). An elementary introduction to the Langlands program. *Bulletin of the American Mathematical Society (N.S.)*, *10*(2), 177–219. 1.4, 41.5.11

[GM2004] Gelbart, S. S., & Miller, S. D. (2004). Riemann's zeta function and beyond. *Bulletin of the American Mathematical Society (N.S.)*, *41*(1), 59–112. 26.3.15

[GG90] Gel'fand, I. M., Graev, M. I., & Pyatetskii-Shapiro, I. I. (1990). *Representation theory and automorphic functions* (K. A. Hirsch, Trans.), Generalized Functions (Vol. 6). Boston: Academic Press. 37.7

[Gib1891] Gibbs, J. W. (1891). On the rôle of quaternions in the algebra of vectors. *Nature*, *43*, 511–513. 1.1

[Gir83] Girard, P. R. (1984). The quaternion group and modern physics. *European Journal of Physics*, *5*, 25–32. 1.5

[GT2010] Giraud, O., & Thas, K. (2010). Hearing the shape of drums: Physical and mathematical aspects of isospectrality. *Reviews of Modern Physics*, *82*, 2213–2255. 31.8

[God58a] Godement, R. (1958–1960). *Les fonctions ζ des algèbres simples, I*, Séminaire N. Bourbaki, exp. no. 171, 27–49. 29.10.24

[God58b] Godement, R. (1958–1960). *Les fonctions ζ des algèbres simples, II*, Séminaire N. Bourbaki, , exp. no. 176, 109–128. 29.10.24

[God62] Godement, R. (1962/63). *Domaines fondamentaux des groupes arithmétiques*, 1964 Séminaire Bourbaki, 1962/63, Fasc. 3, No. 257, Secrétariat mathématique, Paris. 38.5.6

[GJ72] Godement, R., & Jacquet, H. (1972). *Zeta functions of simple algebras*. Lecture Notes in Mathematics (Vol. 260). Berlin: Springer. 29.10.24

[Gol85] Goldfeld, D. (1985). Gauss' class number problem for imaginary quadratic fields. *Bulletin of the American Mathematical Society*, *13*, 23–37. 25.2.19

[Gor2000] Gordon, C. S. (2000). *Survey of isospectral manifolds*, Handbook of differential geometry (Vol. I, pp. 747–778). North-Holland, Amsterdam.

[Gor2009] Gordon, C. S. (2009). *Sunada's isospectrality technique: Two decades later*, Spectral analysis in geometry and number theory, Contemporary Mathematics (Vol. 484, pp. 45–58). Providence, RI: American Mathematical Society.

[Gol71] Goldstein, L. J. (1971). *Analytic number theory*. Englewood Cliffs, N.J.: Prentice-Hall Inc. 27.4

[Gor2012] Gordon, C. S. (2012). *Orbifolds and their spectra*, Spectral geometry, Proceedings of Symposia in Pure Mathematics (Vol. 84, pp. 49–71). Providence, RI: American Mathematical Society. 34.8

[GWW92] Gordon, C., Webb, D. L., & Wolpert, S. (1992). One cannot hear the shape of a drum. *Bulletin of the American Mathematical Society (N.S.)*, *27*(1), 134–138. 31.8

[Gor52] Gorenstein, D. (1952). An arithmetic theory of adjoint plane curves. *Transactions of the American Mathematical Society*, *72*, 414–436. 24.2.24

[Gou97] Gouvêa, F. Q. (1997). *p-adic numbers: An introduction*, Universitext (2nd ed.). Berlin: Springer. 12.1

[Gras1862] Grassmann, H. (1862). *Die, Ausdehnungslehre. Vollstaändig und in strenger Form begründet [Extension theory]*. Berlin: Wiegand. 1.1

[Grav1882] Graves, R. P. (1882). *Life of Sir William Rowan Hamilton*, Volume I, Dublin University Press. 1.1

[Grav1885] Graves, R. P. (1885). *Life of Sir William Rowan Hamilton*, Volume II, Dublin University Press. 1.1, 1.1

[Grav1889] Graves, R. P. (1889). *Life of Sir William Rowan Hamilton*, Volume III, Dublin University Press. 1.1, 1.1

[Gray94] Gray, J. (1994). On the history of the Riemann mapping theorem. *Rendiconti del Circolo Matematico di Palermo, Serie II, Supplemento*, *34*, 47–94. 34.8.3

[Gray2000] Gray, J. (2000). *Linear differential equations and group theory from Riemann to Poincaré* (2nd ed.). Boston: Birkhäuser. 1.4

[Gray2013] Gray, J. (2013). *Henri Poincaré: A scientific biography*. Princeton: Princeton University Press. 1.4

[Gri28] Griffiths, L. W. (1928). Generalized quaternion algebras and the theory of numbers. *American Journal of Mathematics*, *50*(2), 303–314. 1.3

[Gro87] Gross, B. H. (1987). Heights and the special values of L-series, Number theory (Montreal, Quebec, 1985). CMS Conference Proceedings (Vol. 7, pp. 115–187). Providence, RI: American Mathematical Society. 42.3.13

[Gro99] Gross, B. (1999). Algebraic modular forms. *Israel Journal of Mathematics*, *113*, 61–93. 41.2.15

[GrLu2009] Gross, B. H., & Lucianovic, M. W. (2009). On cubic rings and quaternion rings. *Journal of Number Theory*, *129*(6), 1468–1478. 3.15, 22.4, 22.4.3, 22.5.13, 24.2

[Gro85] Grosswald, E. (1985). *Representations of integers as sums of squares*. New York: Springer. 30.2

[GB2008] Grove, L. C., & Benson, C. T. (1985). *Finite reflection groups* (2nd ed.), Graduate Texts in Mathematics (Vol. 99). New York: Springer. 11.5

[GrLe87] Gruber, P. M., & Lekkerkerker, C. G. (1987). *Geometry of numbers* (2nd ed.). New York: North-Holland. 17.5

[GQ2004] Guo, X., & Qin, H. (2004). An embedding theorem for Eichler orders. *Journal of Number Theory*, *107*(2), 207–214. 31.7, 31.7.7

[Grov2002] Grove, L. C. (2002). *Classical groups and geometric algebra*, Graduate Studies in Mathematics (Vol. 39). American Mathematical Society. 4.2, 6.3, 28.8

[Hae84] Haefliger, A. (1984). Groupoides d'holonomie et classifiants, Structure transverse des feuilletages, Toulouse 1982. *Astérisque*, *116*, 70–97. 34.8.15

[HM2006] Hallouin, E., & Maire, C. (2006). Cancellation in totally definite quaternion algebras. *Journal für die reine und angewandte Mathematik*, *595*, 189–213. 20.7

[Ham1843] Hamilton, W. R. (1843). On a new species of imaginary quantities connected with a theory of quaternions. *Proceedings of the Royal Irish Academy*, *2*, 424–434. 1.1

[Ham1844] Hamilton, W. R. (1844). On quaternions, or on a new system of imaginaries in algebra: Copy of a letter from Sir William R. Hamilton to John T. Graves, Esq. on quaternions. *Philosophical Magazine*, **25**, 489–495. 1.2

[Ham1853] Hamilton, W. R. (1853). *Lectures on quaternions*. Cambridge: Cambridge University Press. 1.1

[Ham1866] Hamilton, W. R. (1866). *Elements of quaternions*. Cambridge: Cambridge University Press. 1.1.5, 1.1

[Ham1899] Hamilton's quaternions, review of Elements of quaternions. *Nature*, *387* (1899, August 24). 1.1

[Ham67] Hamilton, W. R. (1967). *The mathematical papers of Sir William Rowan Hamilton, Vol. III: Algebra*, H. Halberstam and R. E. Ingram (Eds.). Cambridge University Press. 1.1, 1.1

[Hanlon81] Hanlon, P. (1981). *Applications of quaternions to the study of imaginary quadratic class groups*, Ph.D. thesis, California Institute of Technology. 30.2

[Hankin80] Hankins, T. L. (1980). *William Rowan Hamilton*. Johns Hopkins University Press. 1.1, 1.1

[Hans2006] Hanson, A. J. (2006). *Visualizing quaternions*. San Francisco: Morgan Kaufmann. 1.5, 2.4.22

[Hap80] Happel, D. (1980). Klassifikationstheorie endlich-dimensionaler Algebren in der Zeit von 1880 bis 1920. *Enseignement Mathématique (2)*, *26*(1–2), 91–102. 1.2

[Har63a] Harada, M. (1963). Hereditary orders. *Transactions of the American Mathematical Society*, *107*(2), 273–290. 20.3, 21.4

[Har63b] Harada, M. (1963). Structure of hereditary orders over local rings. *Journal of Mathematics, Osaka City University*, *14*, 1–22. 21.4

[Har63c] Harada, M. (1963c). Hereditary orders in generalized quaternions D_τ. *Journal of Mathematics, Osaka City University*, *14*, 71–81. 21.4

[Har1881] Hardy, A. S. (1881). *Elements of quaternions*. Boston: Ginn, Heath, and Company.

[HFK94] Hart, J. C., Francis, G. K., & Kauffman, L. H. (1994). Visualizing quaternion rotation. *ACM Transactions on Graphics*, *13*(3), 256–276. 1.5

[Har77] Hartshorne, R. (1977). *Algebraic geometry*, Graduate Texts in Mathematics (Vol. 52). New York: Springer. 4.13, 9.16

[Hash77] Hashimoto, K. (1977). Twisted trace formula of the Brandt matrix. *Proceedings of the Japan Academy, Series A*, *53*, 98–102. 30.6.18

[HM95] Hashimoto, K., & Murabayashi, N. (1995). Shimura curves as intersections of Humbert surfaces and defining equations of QM-curves of genus two. *Tôhoku Mathematical Journal*, *47*, 271–296. 43.2.21

[Hass34] Hasse, H. (1934). Über gewisse Ideale in einer einfachen Algebra. *Act. Sci. Ind. Paris*, *109*, 12–16. 30.3.17

[Hass52] Hasse, H. (1952). *Über die Klassenzahl abelscher Zahlkörper*. Berlin: Akademie-Verlag. 39.4.11

[Hass67] Hasse, H. (1967). History of class field theory. In J. W. S. Cassels & A. Fröhlich (Eds.), *Algebraic number theory* (pp. 266–279). New York: Academic Press. 1.3

[Haw2000] Hawkins, T. (2000). *The emergence of the theory of Lie groups: An essay in the history of mathematics, 1869–1926*. New York: Springer. 1.2

[Hec40] Hecke, E. (1940). Analytische Arithmetik der positiven quadratischen Formen. *Danske Videnskabernes Selskab. Matematisk-fysiske Meddelelser*, *17*(12). 1.4, 41.5.11

[Heeg52] Heegner, K. (1952). Diophantische Analysis und Modulfunktionen. *Mathematische Zeitschrift*, *56*, 227–253. 25.2.19

[Hel2001] Helgason, S. (2001). *Differential geometry, Lie groups, and symmetric spaces*, Graduate Studies in Mathematics (Vol. 34). Providence, RI: American Mathematical Society. 34.6.7

[Herm1854] Hermite, C. (1854). Sur la théorie des formes quadratiques: Premier mémoire. *Journal für die reine und angewandte Mathematik*, *47*, 307–342. 22.5.13

[Hey29] Hey, K. (1929). *Analytische Zahlentheorie in Systemen hyperkomplexer Zahlen*, Ph.D. dissertation, Universität Hamburg. 25.1, 26.8, 29.10.24, 38.4.5

[Hij74] Hijikata, H. (1974). Explicit formula of the traces of Hecke operators for $\Gamma_0(N)$. *Journal of the Mathematical Society of Japan*, *26*(1), 56–82. 23.4, 23.4.22, 30.6, 30.6.18, 41.5.11

[HPS89a] Hijikata, H., Pizer, A. K., & Shemanske, T. R. (1989). The basis problem for modular forms on $\Gamma_0(N)$. *Memoirs of the American Mathematical Society*, **82**(418). 1.4, 24.5.7, 30.6.18, 41.5.11

[HPS89b] Hijikata, H., Pizer, A., & Shemanske, T. (1989). Orders in quaternion algebras. *Journal für die reine und angewandte Mathematik*, *394*, 59–106. 24.5.7, 30.6.18

[HN94] Hijikata, H., & Nishida, K. (1994). Bass orders in non semisimple algebras. *Journal of Mathematics Kyoto University*, *34*(4), 797–837. 21.4, 21.5

[HS73] Hijikata, H., & Saito, H. (1973). *On the representability of modular forms by theta series*, Number Theory, Algebraic Geometry, and Commutative Algebra in honor of Y. Akizuki (pp. 13–21). Kinokuniya, Tokyo. 41.5.11

[Hil32] Hilbert, D. (1932). *Gesammelte Abhandlungen, Bd. 1: Zahlentheorie*. Berlin: Springer. 14.6.4

[Hir53] Hirzebruch, F. (1953). Über vierdimensionale Riemannsche Flächen mehrdeutiger analytischer Funktionen von zwei komplexen Veränderlichen. *Mathematische Annalen*, *126*, 1–22. 35.4

[Hir91] Hirzebruch, F. (1991). *Division algebras and topology*, Numbers, Graduate Texts in Mathematics (Vol. 123, pp. 281–302). New York: Springer. 1.2

[HK2004] Hoehnke, H.-J., & Knus, M.-A. (2004). *Algebras, their Invariants, and K-forms: A tribute to [the work of] HEINRICH BRANDT on the 50th anniversary of his death*, preprint, 2004. https://people.math.ethz.ch/~knus/papers/brandt04.pdf. 1.4

[Hun99] Huneke, C. (1999). *Hyman Bass and ubiquity: Gorenstein rings*, Algebra, K-theory, groups, and education (New York, 1997). Contemporary Mathematics (Vol. 243, pp. 55–78). Providence, RI: American Mathematical Society. 24.2.24

[Hur1898] Hurwitz, A. (1898). *Über die Komposition der quadratischen Formen von beliebig vielen Variablen*, Nach. der köbig. Gesell. Wissen. Göttingen, Math.-Physik. Klasse, 309–316. 1.3

[Hur1896] Hurwitz, A. (1896). Über die Zahlentheorie der Quaternionen. *Nachrichten der Gesellschaft der Wissenschaften zu Göttingen*, 314–340. 1.3

[Hur19] Hurwitz, A. (1919). *Vorlesungen über die Zahlentheorie der Quaternionen*. Berlin: Springer. 1.3, 11.1

[Ibu82] Ibukiyama, T. (1982). On maximal orders of division quaternion algebras over the rational number field with certain optimal embeddings. *Nagoya Mathematical Journal*, *88*, 181–195. 15.5.7

[Igu60] Igusa, J.-I. (1960). Arithmetic variety of moduli for genus two. *Annals of Mathematics*, *72*(3), 612–649. 43.3.5, 43.3.8, 43.5.7

[Ive92] Iversen, B. (1992). *Hyperbolic geometry*, London Mathematical Society Student Texts (Vol. 25). Cambridge: Cambridge University Press. 33.1, 36.1

[Iwa92] Iwasawa, K. (1992). *Letter to J. Dieudonné, dated April 8, 1952*, Zeta functions in geometry, N. Kurokawa and T. Sunada (Eds.), Advanced Studies in Pure Mathematics (Vol. 21, pp. 445–450). 29.10.24

[Jaci68] Jacobinski, H. (1968). Genera and decompositions of lattices over orders. *Acta Mathematica*, *121*, 1–29. 20.7.23

[Jaci71] Jacobinski, H. (1971). Two remarks about hereditary orders. *Proceedings of the American Mathematical Society*, *28*(1), 1–8. 21.2, 21.4

[Jacn74] Jacobson, N. (1974). Abraham Adrian Albert (1905–1972). *Bulletin of the American Mathematical Society*, *80*, 1075–1100. 1.3

[Jacn2009] Jacobson, N. (2009). *Finite-dimensional division algebras over fields*. Berlin: Springer. 7.2, 8.2

[JL70] Jacquet, H., & Langlands, R. (1970). *Automorphic forms on GL(2)*, Lecture Notes in Mathematics (Vol. 114). New York: Springer. 1.4, 41.5.11

[JKS97] Jagy, W. C., Kaplansky, I., & Schiemann, A. (1997). There are 913 regular ternary forms. *Mathematika*, *44*(2), 332–341. 25.4

[Jah2010] Jahangiri, M. (2010). Generators of arithmetic quaternion groups and a Diophantine problem. *International Journal of Number Theory*, *6*(6), 1311–1328. 32.1

[Jan96] Janusz, G. J. (1996). *Algebraic number fields* (2nd ed.), Graduate Studies in Mathematics (Vol. 7). Providence, RI: American Mathematical Society. 14.4, 14.4.4, 14.4.12, 27.5

[JS87] Jones, G. A., & Singerman, D. (1987). *Complex functions: An algebraic and geometric viewpoint*. Cambridge University Press. 33.1, 34.7

[Jun97] Jun, S. (1997). On the certain primitive orders. *Journal of Korean Mathematical Society*, *34*(4), 791–807. 24.5.7

[Jun08] Jung, H. W. E. (1908). Darstellung der Funktionen eines algebraischen Körpers zweier unabhängigen Veränderlichen x, y in der Umgebung einer Stelle $x = a, y = b$. *Journal für die reine und angewandte Mathematik*, *133*, 289–314. 35.4

[Kac66] Kac, M. (1966). Can one hear the shape of a drum?. *American Mathematical Monthly*, **73**(4), part II, 1–23.

[Kap69] Kaplansky, I. (1969). Submodules of quaternion algebras. *Proceedings of the London Mathematical Society (3)*, *19*, 219–232. 16.6, 16.6.9, 16.8, 16.18, 19.5.8

[Kar2010] Karoubi, M. (2010). *K-theory, an elementary introduction*, Cohomology of groups and algebraic K-theory, Advanced Lectures in Mathematics (ALM) (Vol. 12, pp. 197–215). Somerville, MA: International Press. 12.4.4

[Kat85] Katok, S. (1985). Closed geodesics, periods and arithmetic of modular forms. *Inventiones mathematicae*, *80*, 469–480. 43.2.8

[Kat92] Katok, S. (1992). *Fuchsian groups*. Chicago: University of Chicago Press. 33.1, 33.6, 37.3, 37.7

[Kat2007] Katok, S. (2007). *p-adic analysis compared with real*, Student Mathematical Library (Vol. 37). Providence, RI: American Mathematical

Society. 12.1

[Ker58] Kervaire, M. (1958). Non-parallelizability of the n-sphere for $n>7$. *Proceedings of the National Academy of Sciences*, *44*, 280–283. 1.2

[Kil1888] Killing, W. (1888). Die Zusammensetzung der stetigen endlichen Transformations-gruppen. *Mathematische Annalen*, **31**, 252–290; *Mathematische Annalen*, *33*, 1–48; *Mathematische Annalen*, *34*, 57–122 (1889); *Mathematische Annalen*, *36*(2), 161–189 (1890). 1.2

[Kir2005] Kirschmer, M. (2005). *Konstruktive Idealtheorie in Quaternionenalgebren*, Ph.D. thesis, Universität Ulm. 17.7.25

[Kir2014] Kirschmer, M. (2014). One-class genera of maximal integral quadratic forms. *Journal of Number Theory*, *136*, 375–393. 25.4.7

[KL2016] Kirschmer, M., & Lorch, D. (2016). Ternary quadratic forms over number fields with small class number. *Journal of Number Theory*, *161*, 343–361. 26.7, 26.7.4

[KLwww] Kirschmer, M., & Lorch, D. *The one-class genera of ternary quadratic forms*, website http://www.math.rwth-aachen.de/~kirschme/orders/. 26.7.4

[Kle1878] Klein, F. (1878–1879). Über die Transformation der elliptischen Funktionen und die Auflösung der Gleichungen fünften Grades. *Mathematische Annalen*, **14**, 111–172. 1.4

[Kle56] Klein, F. (1956). *Lectures on the icosahedron and the solution of equations of the fifth degree*, translated by George Gavin Morrice (2nd revised ed.). New York: Dover Publications. 32.4

[Kle79] Klein F. (1979). *Development of mathematics in the 19th century*, translated by M. Ackerman, Lie Groups: History, frontiers and applications (Vol. IX). Brookline, Massachusetts: Math Sci Press. 34.3.10

[Klt2000] Kleinert, E. (2000). *Units in skew fields*, Progress in Mathematics (Vol. 186). Basel: Birkhäuser Verlag. 28.8, 38.4

[Kna2016] Knapp, A. W. (2016). *Advanced algebra*, digital 2nd edition, East Setauket, New York. 27.4

[Kne66a] Kneser, M. (1965). Starke approximation in algebraischen gruppen. I. *Journal für die reine und angewandte Mathematik*, *218*, 190–203. 28.7.10, 28.8

[Kne66b] Kneser, M. (1966). *Strong approximation*, Algebraic Groups and Discontinuous Subgroups (Boulder, Colorado, 1965). Proceedings of Symposia in Pure Mathematics (pp. 187–196). Providence, RI: American Mathematical Society. 28.7.10, 28.8

[KKOPS86] Kneser, M., Knus, M.-A., Ojanguren, M., Parimala, R., & Sridharan, R. (1986). Composition of quaternary quadratic forms. *Compositio Mathematica*, *60*(2), 133–150. 19.5.8

[Knu88] Knus, M.-A. (1988). *Quadratic forms, Clifford algebras and spinors*, Seminários de Matemática, 1. Universidade Estadual de Campinas, Instituto de Matemática, Estatística e Ciência da Computaç ã o, Campinas. 9.7

[Knu93] Knus, M.-A. (1993). Sur la forme d'Albert et le produit tensoriel de deux algebres de quaternions. *Bulletin of the Belgian Mathematical Society*, *45*, 333–337. 8.2.6

[KMRT98] Knus, M.-A., Merkurjev, A., Rost, M., & Tignol, J.-P. (1998). *The book of involutions*, American Mathematical Society Colloquium Publications (Vol. 44). Providence, RI: American Mathematical Society. 4.4.6, 5.6, 5.6.11

[Kob84] Koblitz, N. (1984). *p-adic numbers, p-adic analysis, and zeta-functions*, Graduate Texts in Mathematics (Vol. 58, 2nd ed.). New York: Springer. 12.1

[Koh96] Kohel, D. (1996). *Endomorphism rings of elliptic curves over finite fields*. Berkeley: University of California. Ph.D. thesis. 42.3.6

[Koh2001] Kohel, D. (2001). *Hecke module structure of quaternions*, Class field theory: Its centenary and prospect (Tokyo, 1998). K. Miyake (Ed.), Advanced Studies in Pure Mathematics (Vol. 30, pp. 177–195). Tokyo: Mathematical Society of Japan. 41.5.12

[KV2003] Kohel, D. R., & Verrill, H. A. (2003). Fundamental domains for Shimura curves. *Journal de Théorie des Nombres de Bordeaux*, *15*, 205–222. 37.9.13

[Kör85] Körner, O. (1985). Über die zentrale Picard-Gruppe und die Einheiten lokaler Quaternionenordnungen. *Manuscripta Mathematica*, *52*, 203–225. 26.6

[Kör87] Körner, O. (1987). Traces of Eichler–Brandt matrices and type numbers of quaternion orders. *Proceedings of the Indian Academy of Sciences (Mathematical Sciences)*, *97*(1–3), 189–199. 25.3, 26.1.14, 30.8.8, 30.9.12

[Kur79] Kurihara, A. (1979). On some examples of equations defining Shimura curves and the Mumford uniformization. *Journal of the Faculty of Science, University of Tokyo. Section IA: Mathematics*, *25*(3), 277–300. 43.2

[Lam99] Lam, T.-Y. (1999) *Lectures on modules and rings*, Graduate Texts in Mathematics (Vol. 189). New York: Springer. 7.2.20, 20.2.2, 20.3.6

[Lam2001] Lam, T.-Y. (2001). *A first course in noncommutative rings*, Graduate Texts in Mathematics (Vol. 131, 2nd ed.). New York: Springer. 7.2

[Lam2002] Lam, T.-Y. (2002). On the linkage of quaternion algebras. *Bulletin of the Belgian Mathematical Society*, *9*, 415–418. 8.2.6

[Lam2003] Lam, T.-Y. (2003). *Hamilton's quaternions*, Handbook of algebra (Vol. 3, pp. 429–454). North Holland, Amsterdam. 1

[Lam2005] Lam, T.-Y. (2005). *Introduction to quadratic forms over fields*, Graduate Studies in Mathematics (Vol. 67). Providence, RI: American Mathematical Society. 3.5.3, 4.2, 4.2.22, 4.5, 5.2, 5.4, 8.2.9, 26.8

[Lam2006] Lam, T.-Y. (2006) *Serre's problem on projective modules*, Springer Monographs in Mathematics. Berlin: Springer. 20.7.24

[Lamo86] Lamotke, K. (1986). *Regular solids and isolated singularities*, Advanced Lectures in Mathematics. Braunschweig: Friedr. Vieweg and Sohn. 32.4

[Lanc67] Lanczos, C. (1967). William Rowan Hamilton–An appreciation. *American Scientist*, *55*, 129–143. 1.1

[Lang94] Lang, S. (1994). *Algebraic number theory*, Graduate Texts in Mathematics, (Vol. 110, 2nd ed.). Berlin: Springer. 26.2, 26.2.18, 26.8.26, 27.4, 27.5

[Lang95] Lang, S. (1995). *Introduction to modular forms*, appendixes by D. Zagier and Walter Feit, Grundlehren der Mathematischen Wissenschaften (Vol. 222). Berlin: Springer. 40.1

[Lang82] Lang, S. (1982). *Introduction to algebraic and abelian functions*, Graduate Texts in Mathematics (Vol. 89, 2nd ed.). New York: Springer. 43.6

[Lat26] Latimer, C. G. (1926). Arithmetics of generalized quaternion algebras. *American Journal of Mathematics (2)*, *27*, 92–102. 1.3

[Lat35] Latimer, C. G. (1935). On the fundamental number of a rational generalized quaternion algebra. *Duke Mathematical Journal*, *1*(4), 433–435. 14.2.8

[Lat37] Latimer, C. G. (1937). The classes of integral sets in a quaternion algebra. *Duke Mathematical Journal*, *3*, 237–247. 22.5.13

[Lee2011] Lee, J. M. (2011). *Introduction to topological manifolds*, Graduate Texts in Mathematics (Vol. 202, 2nd ed.). New York: Springer. 34.5

[Lem2011] Lemurell, S. (2011). *Quaternion orders and ternary quadratic forms*. arXiv:1103.4922. 16.12, 22.5.13, 24.1.5, 24.6

[Len96] Lenstra, Jr., H. W. (1996). Complex multiplication structure of elliptic curves. *Journal of Number Theory*, *56*, 227–241. 42.1

[LS91] Lenstra, H. W., & Stevenhagen, P. (1991). Primes of degree one and algebraic cases of Cebotarev's theorem. *Enseignement mathématique*, *37*, 17–30. 14.2.10

[Lev2013] Levin, A. (2013). *On the classification of algebras*, M.Sc. thesis, University of Vermont. 3.5.11

[LR2011] Levy, L. S., & Robson, J. C. (2011). *Hereditary Noetherian prime rings and idealizers*, Mathematical Surveys and Monographs (Vol. 174). Providence, RI: American Mathematical Society. 21.4.11

[Lew2006a] Lewis, D. W. (2006). Quaternion algebras and the algebraic legacy of Hamilton's quaternions. *Irish Mathematical Society Bulletin*, *57*, 41–64. 1

[Lew2006b] Lewis, D. W. (2006). Involutions and anti-automorphisms of algebras. *Bulletin of the London Mathematical Society*, *38*, 529–545. 1.2

[LL2012] Li, W.-C. W., & Long, L. (2012). *Atkin and Swinnerton-Dyer congruences and noncongruence modular forms*, Algebraic number theory and related topics 2012, RIMS Kôkyûroku Bessatsu, B51, Research Institute for Mathematical Sciences (RIMS), Kyoto, 2014 (pp. 269–299). arXiv:1303.6228. 35.4.4

[Lin2012] Linowitz, B. (2012). Selectivity in quaternion algebras. *Journal of Number Theory*, *132*(7), 1425–1437. 31.7.7

[LS2012] Linowitz, B., & Shemanske, T. R. (2012). Embedding orders into central simple algebras. *Journal de Théorie des Nombres de Bordeaux*, *24*(2), 405–424. 31.7.7

[LV2015] Linowitz, B., & Voight, J. (2015). Small isospectral and nonisometric orbifolds of dimension 2 and 3. *Mathematische Zeitschrift*, *281*(1–2), 523–569.

[Liv2001] Livné, R. (2001). *Communication networks and Hilbert modular forms*, Applications of algebraic geometry to coding theory, physics and computation (Eilat, 2001). NATO Science Series II Mathematics, Physics and Chemistry (Vol. 36, pp. 255–270). Dordrecht: Kluwer Academic Publishers. 41.2.14

[LP2010] Lobachevsky, N. I. (2010). *Pangeometry*, A. Papadopoulos (Ed.). European Mathematical Society. 33.1

[Loo53] Loomis, L. H. (1953). *An introduction to abstract harmonic analysis*. Princeton: D. Van Nostrand Company Inc. 29.3

[LK2013] Lorch, D., & Kirschmer, M. (2013). Single-class genera of positive integral lattices. *LMS Journal of Computation and Mathematics*, *16*, 172–186. 25.4, 25.4.7

[Lub2010] Lubotzky, A. (2010). *Discrete groups, expanding graphs and invariant measures*, appendix by Jonathan D. Rogawski, reprint edition. Basel: Birkhäuser Verlag. 41.2.14

[LPS88] Lubotzky, A., Phillips, R., & Sarnak, P. (1988). Ramanujan graphs. *Combinatorica*, *8*, 261–277. 41.2.14

[Luc2003] Lucianovic, M. (2003). *Quaternion rings, ternary quadratic forms, and Fourier coefficients of modular forms on* PGSp(6), Ph.D. thesis, Harvard University. 22.5.13

[Macf1891] Macfarlane, A. (1891). Principles of the algebra of physics. *Proceedings of the American Association for the Advancement of Science*, *15*, 65–112. 1.2

[Macf00] Macfarlane, A. (1900). Hyperbolic quaternions. *Proceedings of the Royal Society of Edinburgh*, *23*, 169–180. 1.2

[Macl2008] Maclachlan, C. (2008). Optimal embeddings in quaternion algebras. *Journal of Number Theory*, *128*, 2852–2860. 31.1.8, 31.6

[MR2003] Maclachlan, C., & Reid, A. W. (2003). *The arithmetic of hyperbolic 3-manifolds*, Graduate Texts in Mathematics (Vol. 219). New York: Springer.

[MacRos94] Maclachlan, C., & Rosenberger, G. (1994). Small volume isospectral, nonisometric, hyperbolic 2 -orbifolds. *Archiv der Mathematik (Basel)*, *62*(1), 33–37. 31.8

[Mard2007] Marden, A. (2007). *Outer circles: An introduction to hyperbolic 3-manifolds*. Cambridge: Cambridge University Press. 36.1

[Marg88] Margulis, G. (1988). Explicit group theoretic constructions of combinatorial schemes and their application to the design of expanders and concentrators. *Journal Problems of Information Transmission*, 39–46. 41.2.14

[Mas88] Maskit, B. (1988). *Kleinian groups*. Springer. 37.6.5

[Mas71] Maskit, B. (1971). On Poincaré's theorem for fundamental polygons. *Advances in Mathematics*, *7*, 219–230. 37.6.4, 37.6.5

[Mat46] Matchett, M. (1946). *On the zeta function for ideles*, Ph.D. thesis, Indiana University. 29.10.24

[Mat89] Matsumura, H. (1989). *Commutative ring theory*, translated by M. Reid, Cambridge Studies in Advanced Mathematics (Vol. 8, 2nd ed.). Cambridge: Cambridge University Press. 9.2, 22.2

[Mat69] Matsumoto, H. (1969). Sur les sous-groupes arithmétiques des groupes semi-simples déployés. *Annales Scientifiques de l'École Normale Supérieure (4)*, *2*, 1–62. 12.4.4

[Max1869] Maxwell, J. C. (1869). Remarks on the mathematical classification of physical quantities. *Proceedings of the London Mathematical Society*, *3*, 224–232. 1.1

[May66] May, K. O. (1966). The impossibility of a division algebra of vectors in three dimensional space. *American Mathematical Monthly*, *73*(3), 289–291. 1.1

[McC2011] McCarty, G. (2010). *Topology: An introduction with application to topological groups*. New York: Dover. 34.1, 34.3.11

[McCR87] McConnell, J. C., & Robson, J. C. (1987). *Noncommutative Noetherian rings*, Graduate Students in Mathematics (Vol. 30). Providence, RI: American Mathematical Society. 18.3.9

[Mer82] Merkurjev, A. (1982). On the norm residue symbol of degree 2. *Soviet Mathematics Doklady*, *24*, 1546–1551. 8.3.6

[Mes91] Mestre, J.-F. (1991). *Construction de courbes de genre 2 à partir de leurs modules*, Effective methods in algebraic geometry, Progress in Mathematics (Vol. 94, pp. 313–334). Boston: Birkhäuser. 43.3.13

[Milne80] Milne, J. S. (1980). *Étale cohomology*, Princeton Mathematical Series (Vol. 33). Princeton: Princeton University Press. 15.17

[Milne2017] Milne, J. S. (2017). *Algebraic groups: The theory of group schemes of finite type over a field*, Cambridge Studies in Advanced Mathematics (Vol. 170). Cambridge: Cambridge University Press. 38.5

[Milne-CFT] Milne, J. S. *Class field theory*. http://www.jmilne.org/math/CourseNotes/cft.html. 14.6.10, 26.8

[Milne-SV] Milne, J. S. *Introduction to Shimura varieties*. https://www.jmilne.org/math/xnotes/svi.html. 38.7.20

[Mil64] Milnor, J. (1964). Eigenvalues of the Laplace operator on certain manifolds. *Proceedings of the National Academy of Sciences of the United States of America*, *51*, 542.

[Milno71] Milnor, J. (1971). *Introduction to algebraic K-theory*, Annals of Mathematics Studies, no. 72. Princeton: Princeton University Press and University of Tokyo Press. 12.4.4

[Milno82] Milnor, J. (1982). Hyperbolic geometry: The first 150 years. *Bulletin of the American Mathematical Society*, *6*(1), 9–24. 33.1, 36.5, 36.5.13

[Mir95] Miranda, R. (1995). *Algebraic curves and Riemann surfaces*, Graduate Studies in Mathematics (Vol. 5). Providence, RI: American Mathematical Society. 33.8.6

[Miy2006] Miyake, T. (2006). *Modular forms*, Translated from the 1976 Japanese original by Yoshitaka Maeda, Springer Monographs in Mathematics. Berlin: Springer. 28.7, 40.1, 40.4, 40.5

[Moe2002] Moerdijk, I. (2002). *Orbifolds as groupoids: An introduction*, Orbifolds in mathematics and physics (Madison, WI, 2001). Contemporary Mathematics (Vol. 310, pp. 205–222). Providence, RI: American Mathematical Society. 34.8.15

[MP1997] Moerdijk, I., & Pronk, D. A. (1997). Orbifolds, sheaves and groupoids. *K-Theory*, *12*(1), 3–21. 34.8.15

[Mol1893] Molien, T. (1893). Über Systeme höherer complexer Zahlen. *Mathematische Annalen*, *41*, 83–156; Berichtigung zu dem Aufsatze 'Über Systeme höherer complexer Zahlen'. *Mathematische Annalen*, *42*, 308–312. 1.2

[MW77] Montgomery, H. L., & Weinberger, P. J. (1977). Real quadratic fields with large class number. *Mathematische Annalen*, *225*(2), 173–176. 26.7.1

[Moore35] Moore, E. H. (1935). *General analysis*, Part I, Memoirs of the American Philosophical Society (Vol. 1). Philadelphia: American Philosophical Society. 3.5.3

[Mor69] Mordell, L. J. (1969). *Diophantine equations*, Pure Applied Mathematics (Vol. 30). London: Academic Press. 11.4.4

[MT62] Mostow, G. D., & Tamagawa, T. (1962). On the compactness of arithmetically defined homogeneous spaces. *Annals of Mathematics (2)*, **76**, 446–463. 38.5.6

[Mum70] Mumford, D. (1970). *Abelian varieties*, Tata Institute of Fundamental Research Studies in Mathematics, no. 5. Oxford: Oxford University Press. 8.5, 43.4

[Mun2004] Munn, W. D. (2004). Involutions on finite-dimensional algebras over real closed fields. *Journal of the Australian Mathematical Society*, *77*, 123–128. 8.4.17

[Mur88] Murty, M. R. (1988). Primes in certain arithmetic progressions. *Journal of Madras University, Section B*, *51*, 161–169. 14.2.10

[Neu99] Neukirch, J. (1999). *Algebraic number theory*, Grundlehren der mathematischen Wissenschaften (Vol. 322). Berlin: Springer. 12.2, 12.2.15, 13.2, 14.2.10, 14.4, 14.4.4, 14.4.12, 15.1, 26.2.18, 26.2, 26.8, 27.4, 27.5, 28.7.1, 28.7.2

[New72] Newman, M. (1972). *Integral matrices*, Pure and Applied Mathematics (Vol. 45). New York: Academic Press. 17.2

[NSW2008] Neukirch, J., Schmidt, A., & Wingberg, K. (2008). *Cohomology of number fields*, Grundlehren der mathematischen Wissenschaften (Vol. 323, 2nd ed.) Berlin: Springer. 14.6.10

[Nip74] Nipp, G. L. (1974). Quaternion orders associate with ternary lattices. *Pacific Journal of Mathematics*, *53*(2), 525–537. 22.5.13

[Noe34] Noether, E. (1934). Zerfallende verschränkte Produkte und ihre Maximalordnungen. *Act. Sci. Ind. Paris*, *148*, 15pp. 30.3.17

[ÓCa2000] Ó Cairbre, F. (2000). William Rowan Hamilton (1805–1865), Ireland's greatest mathematician. *Ríocht na Midhe* (Meath Archaeological and Historical Society), *11*, 124–150. 1.1

[ÓCa2010] Ó Cairbre, F. (2010). Twenty years of the Hamilton walk. *Irish Mathematical Society Bulletin*, *65*, 33–49. 1.1

[O'Do83] O'Donnell, S. (1983). *William Rowan Hamilton*. Boole Press Limited. 1.1

[Ogg69] Ogg, A. (1969). *Modular forms and Dirichlet series*. New York: W.A. Benjamin Inc. 40.4

[Ogg83] Ogg, A. (1983). *Real points on Shimura curves*, Arithmetic and geometry, Vol. I, Progress in Mathematics (Vol. 35, pp. 277–307). Boston: Birkhäuser Boston. 43.7

[O'Me73] O'Meara, O. T. (1973). *Introduction to quadratic forms*. Berlin: Springer. 4.2, 4.2.22, 4.5, 9.3, 9.3.6, 9.7, 14.6, 28.7.1, 28.7.2

[PS2014] Pacetti, A., & Sirolli, N. (2014). Computing ideal classes representatives in quaternion algebras. *Mathematics of Computation*, *83*(289), 2479–2507. 24.6, 26.6.1

[Pall46] Pall, G. (1946). On generalized quaternions. *Transactions of the American Mathematical Society*, *59*, 280–332. 22.5.13

[Pap2014] Papadopoulos, A. (2014). *Metric spaces, convexity and non-positive curvature*, IRMA Lectures in Mathematics and Theoretical Physics (Vol. 6, 2nd ed.). Zürich: European Mathematical Society. 33.2.8

[PS2018] Parzanchevski, O., & Sarnak, P. (2018). Super-golden-gates for $PU(2)$. *Advances in Mathematics*, *327*, 869–901. 1.5

[Pei1882] Peirce, B. (1881). Linear associative algebra. *American Journal of Mathematics*, *4*(1–4), 97–229. 1.2

[Pet69] Peters, M. (1969). Ternäre und quaternäre quadratische Formen und Quaternionenalgebren. *Acta Arithmetica*, *15*, 329–365. 22.5.13, 30.9.12

[Pie82] Pierce, R. S. (1982). *Associative algebras*. New York: Springer. 7.2, 7.9

[Piz73] Pizer, A. K. (1973). Type numbers of Eichler orders. *Journal für die reine und angewandte Mathematik*, *264*, 76–102. 23.4.22, 30.9.12

[Piz76a] Pizer, A. (1976). On the arithmetic of quaternion algebras. *Acta Arithmetica*, *31*(1), 61–89. 30.6.18, 30.8.8, 30.9.12

[Piz76b] Pizer, A. (1976). On the arithmetic of quaternion algebras. II. *Journal of the Mathematical Society of Japan*, *28*(4), 676–688. 24.3.7, 30.6.18, 41.5.11

[Piz76c] Pizer, A. (1976). The representability of modular forms by theta series. *Journal of the Mathematical Society of Japan*, *28*, 689–698. 30.6.18, 41.5.11

[Piz80a] Pizer, A. (1980). An algorithm for computing modular forms on $\Gamma_0(N)$. *Journal of Algebra*, *64*, 340–390. 15.5.7, 24.3.7, 41.5.12

[Piz80b] Pizer, A. (1980). Theta series and modular forms of level p^2M. *Compositio Mathematica*, *40*, 177–241. 30.6.18

[Pla69] Platonov, V. P. (1969). Strong approximation in algebraic groups and the Kneser-Tits conjecture. *Doklady Akademii Nauk BSSR*, *13*(7), 585–587. 28.7.10

[Pla69-70] Platonov, V. P. (1969–1970). The strong approximation problem and the Kneser-Tits conjecture for algebraic groups. *Izvestiya Akademii Nauk SSSR, Ser. Math.*, *33*(6), 1211–1219 (1969); Additions to 'The strong approximation problem and the Kneser-Tits conjecture for algebraic groups'. *Izvestiya Akademii Nauk SSSR, Ser. Math.*, *33*(4), 775–777 (1970). 28.7.10, 28.8

[PR94] Platonov, V., & Rapinchuk, A. (1994). *Algebraic groups and number theory*, Pure and Applied Mathematics (Vol. 139). Boston, MA: Academic Press Inc. 28.7.10, 38.5.6

[Pl83] Plesken, W. (1983). *Group rings of finite groups over p-adic integers*, Lecture Notes in Mathematics, no. 1026. Springer. 23.4.23

[Poi1882] Poincaré, H. (1882). Théorie des groups fuchsiens. *Acta Mathematica*, *1*, 1–62. 37.6.5

[Poi1883] Poincaré, H. (1883). Mémoire sur les groupes Kleinéens. *Acta Mathematica*, *3*, 49–92. 37.6.5

[Poi1908] Poincaré, H. (1908). *Science et méthode*. Paris: Flammarion. 1.4

[Pol2010] Pollack, P. (2010). Hypothesis *H* and an impossibility theorem of Ram Murty. *Rendiconti del Seminario Matematico Universitá e Politecnico di Torino*, *68*(2), 183–197. 14.2.10

[Pra77] Prasad, G. (1977). Strong approximation for semi-simple groups over function fields. *Annals of Mathematics (2)*, *105*(3), 553–572. 28.7.10, 28.8

[Pre68] Prestel, A. (1968). Die elliptischen Fixpunkte der Hilbertschen Modulgruppen. *Mathematische Annalen*, *177*, 181–209. 30.6.18

[Puj2012] Pujol, J. (2012). Hamilton, Rodrigues, Gauss, quaternions, and rotations: A historical reassessment. *Communications in Mathematical Analysis*, *13*(2), 1–14. 1.1

[Puj2014] Pujol, J. (2014, June–July). On Hamilton's nearly-forgotten early work on the relation between rotations and quaternions and on the composition of rotations. *American Mathematical Monthly*, *121*, 515–522. 1.1

[Pyn2006] Pynchon, T. (2006). *Against the day*. Penguin Press. 1.1

[Rag72] Raghunathan, M. S. (1972). *Discrete subgroups of Lie groups*, Ergebnisse der Mathematik und ihrer Grenzgebiete (Vol. 68). New York-Heidelberg: Springer. 38.4.7

[Raj93] Rajwade, A. R. (1993). *Squares*, Lecture Note Series (Vol. 171). Cambridge: Cambridge University Press. 3.5.10

[RM99] Ramakrishnan, D., & Valenza, R. J. (1999). *Fourier analysis on number fields*, Graduate Texts in Mathematics (Vol. 186). New York: Springer. 27.4, 29.3

[Ran2011] Ranicki, A. (2011). Commentary on "On the parallelizability of the spheres" by R. Bott and J. Milnor and "On the nonexistence of elements of Hopf invariant one" by J. F. Adams. *Bulletin of the American Mathematical Society (N.S.)*, **48**(4), 509–511. 1.2

[Rap2014] Rapinchuk, A. S. (2014). *Strong approximation for algebraic groups*, Thin groups and superstrong approximation, Mathematical Sciences Research Institute Publications (Vol. 61, pp. 269–298). Cambridge: Cambridge University Press. 28.7.10

[Rat2006] Ratcliffe, J. G. (2006). *Foundations of hyperbolic manifolds*, Graduate Texts in Mathematics (Vol. 149, 2nd ed.). New York: Springer. 34.8, 36.1, 36.3.17, 36.5, 36.5, 37.6

[Reh76] Rehm, H. P. (1976). *On a theorem of Gauss concerning the number of integral solutions of the equation* $x^2 + y^2 + z^2 = m$, Selected topics on ternary forms and norms (Sem. Number Theory, California Institute of Technology, Pasadena, California, 1974/75), Paper No. 4, California Institute of Technology, Pasadena, California, 11pp. 30.2

[Rei70] Reiner, I. (1970). A survey of integral representation theory. *Bulletin of the American Mathematical Society*, *76*(2), 159–227. 20.1.4

[Rei76] Reiner, I. (1976). *Class groups and Picard groups of group rings and orders*, Conference Board of the Mathematical Sciences, Regional Conference Series in Mathematics, no. 26. Providence, RI: American Mathematical Society. 20.1.4

[Rei2003] Reiner, I. (2003). *Maximal orders*. London Mathematical Society Monographs (N.S.) (Vol. 28). Oxford: Clarendon Press, Oxford University Press. 7.2.20, 7.5.5, 7.8.1, 7.8.12, 7.9, 9.2, 10.2, 13.4.3, 15.2, 16.3, 17.7.27, 17.8, 17.12, 18.2, 18.3, 18.3.1, 18.3.2, 20.4, 20.5, 20.5.5, 20.6, 20.7, 21.2.8, 21.4, 21.5.3, 21.6, 21.8, 23.3, 28.6.11

[Rib1989] Ribet, K. A. (1989). Bimodules and abelian surfaces, Algebraic number theory: In honor of K. Iwasawa. *Advances in Pure Mathematics*, *17*, 359–407. 42.4

[Roq2006] Roquette, P. J. (2006). *The Brauer-Hasse-Noether theorem in historical perspective*, Schr. Math.-Nat.wiss. Kl. Berlin-Heidelberg: Springer. 1.3, 29.10.24

[Rod1840] Rodrigues, O. (1840). Des lois geometriques qui regissent les déplacements d'un système solide dans l'espace, et la variation des coordonnées provenant de ses déplacements considérés indépendamment des causes qui peuvent les produire. *Journal de Mathematiques Pures et Appliqués*, *5*, 380–440. 1.1

[Ros2002] Rosen, M. (2002). *Number theory in function fields*, Graduate Texts in Mathematics (Vol. 210). New York: Springer. 29.13

[Rot2004] Rotger, V. (2004). The field of moduli of quaternionic multiplication on abelian varieties. *International Journal of Mathematics and Mathematical Sciences*, *49–52*, 2795–2808. 43.6

[Rud87] Rudin, W. (1987). *Real and complex analysis* (3rd ed.). New York: McGraw-Hill. 29.3.2

[Sah72] Sah, C.-H. (1972). Symmetric bilinear forms and quadratic forms. *Journal of Algebra*, *20*, 144–160. 8.2.6

[Sal99] Saltman, D. J. (1999). *Lectures on division algebras*, CBMS regional conference series in mathematics (Vol. 94). Providence, RI: American Mathematical Society. 7.5, 7.5.6

[San2017] Sands, J. (2017). Zeta-functions and ideal classes of quaternion orders. *Rocky Mountain Journal of Mathematics*, *47*(4), 1277–1300. 26.5.22

[Sar90] Sarnak, P. (1990). *Some applications of modular forms*, Cambridge Tracts in Mathematics (Vol. 99). Cambridge University Press. 41.2.14

[Scha85] Scharlau, W. (1985). *Quadratic and Hermitian forms*. Berlin: Springer. 4.2, 4.2.22, 4.5, 9.7, 9.8

[Scha2009] Scharlau, R. (2009). *Martin Kneser's work on quadratic forms and algebraic groups*, Quadratic forms: Algebra, arithmetic, and geometry, Contemporary Mathematics (Vol. 493, pp. 339–357). Providence, RI: American Mathematical Society. 1.3, 28.7.10, 29.11.9

[Schi35] Schilling, O. (1935). Über gewisse Beziehungen zwischen der Arithmetik hyperkomplexer Zahlsysteme und algebraischer Zahlkörper. *Mathematische Annalen*, *111*(1), 372–398. 30.3.17

[Schn75] Schneider, V. (1975). Die elliptischen Fixpunkte zu Modulgruppen in Quaternionenschiefkörpern. *Mathematische Annalen*, *217*(1), 29–45. 30.6.18

[Schn77] Schneider, V. (1977). Elliptische Fixpunkte und Drehfaktoren zur Modulgruppe in Quaternionenschiefkörpern über reellquadratischen Zahlkörpern. *Mathematische Zeitschrif*, *152*(2), 145–163. 30.6.18

[Sco83] Scott, P. (1983). The geometries of 3-manifolds. *Bulletin of the London Mathematical Society*, *15*, 401–487. 33.1, 34.8

[Ser73] Serre, J.-P. (1973). *A course in arithmetic*, Graduate Texts in Mathematics (Vol. 7). New York: Springer. 14.1, 14.2, 14.3, 25.2.19, 35, 40.1, 40.3

[Ser79] Serre, J.-P. (1979). *Local fields*, Graduate Texts in Mathematics (Vol. 67). New York: Springer. 8.3.7, 12.2, 12.2.15, 13.2, 13.4.7, 38.5

[Ser96] Serre, J.-P. (1996). Two letters on quaternions and modular forms (mod p). *Israel Journal of Mathematics*, *95*, 281–299. 42.3.14

[Ser2003] Serre, J.-P. (2003). *Trees*, Translated by John Stillwell, corrected 2nd printing, Springer Monographs in Mathematics. Berlin: Springer. 23.5

[Sha90] Shapiro, D. B. (2000). *Compositions of quadratic forms*, de Gruyter Expositions in Mathematics (Vol. 33). Berlin: Walter de Gruyter. 1.3

[Shem86] Shemanske, T. R. (1986). Representations of ternary quadratic forms and the class number of imaginary quadratic fields. *Pacific Journal of Mathematics*, *122*(1), 223–250. 30.1

[Shz63] Shimizu, H. (1963). On discontinuous groups operating on the product of the upper half planes. *Annals of Mathematics (2)*, *77*(1), 33–71. 30.6.18

[Shz65] Shimizu, H. (1965). On zeta functions of quaternion algebras. *Annals of Mathematics (2)*, *81*, 166–193. 1.4, 30.6.18, 39.1

[Shz72] Shimizu, H. (1972). Theta series and automorphic forms on GL_2. *Journal of the Mathematical Society of Japan*, *24*, 638–683. 41.5.11

[Shi67] Shimura, G. (1967). Construction of class fields and zeta functions of algebraic curves. *Annals of Mathematics (2)*, *85*, 58–159. 38.7.20, 43.1.9, 43.8.1, 43.8.2

[Shi71] Shimura, G. (1971). *Introduction to the arithmetic theory of automorphic functions*, Kanô Memorial Lectures, no. 1, Publications of the Mathematical Society of Japan, no. 11, Iwanami Shoten, Tokyo, Princeton University Press, Princeton. 28.2, 43.9.4, 43.9.5

[Shi75] Shimura, G. (1975). On the real points of an arithmetic quotient of a bounded symmetric domain. *Mathematische Annalen*, *215*, 135–164. 43.7

[Shi80] Shimura, G. (1980). On some problems of algebraicity. In *Proceedings of the International Congress of Mathematicians* (Helsinki, 1978), Academy of Science Fennica, Helsinki (pp. 373–379). 1.4

[Shi96] Shimura, G. (1996). In 1996 Steele Prizes. *Notices of the American Mathematical Society*, *43*(11), 1340–1347.

[Sho85] Shoemake, K. (1985). Animating rotation with quaternion curves. *ACM SIGGRAPH Computer Graphics*, *19*, 245–254. 1.5

[Sie45] Siegel, C. L. (1945). Some remarks on discontinuous groups. *Annals of Mathematics (2)*, **46**, 708–718. 37.7

[Sie65] Siegel. C. L. (1965). *Vorlesungen über ausgewählte Kapitel der Functionentheorie* (Vol. 2). Lecture notes. Universität Göttingen, Germany. 37.6.5

[Sie69] Siegel, C. L. (1969). Berechnung von Zetafunktionen an ganzzahligen Stellen. *Nachrichten der Akademie der Wissenschaften in Göttingen Mathematisch - Physikalische Klasse II*, 87–102. 26.1.9

[Sie89] Siegel, C. L. (1989). *Lectures on the geometry of numbers*. Berlin: Springer. 17.5

[Sil2009] Silverman, J. H. (2009). *The arithmetic of elliptic curves*, Graduate Texts in Mathematics (Vol. 106, 2nd ed.). New York, Springer. 29.13, 40.1, 40.3.10,42.1,42.1.6,42.1.7,42.1.9,42.1.11,42.2.4, 42.2.7, 42.4.3, 43.3.1

[Sim2010] Simons, P. (2010). Vectors and beyond: Geometric algebra and its philosophical significance. *Dialectica*, *63*(4), 381–395. 1.1

[Sme2015] Smertnig, D. (2015). A note on cancellation in totally definite quaternion algebras. *Journal fü die reine und angewandte Mathematik*, *707*, 209–216. 20.7, 20.7.24

[SV2019] Smertnig, D., & Voight, J. (2019). Definite orders with locally free cancellation. *Transactions of the London Mathematical Society*, *6*(1), 53–86. 20.7, 20.7.24

[SW2005] Socrates, J., & Whitehouse, D. (2005). Unramified Hilbert modular forms, with examples relating to elliptic curves. *Pacific Journal of Mathematics*, *219*(2), 333–364. 41.5.12

[Sol77] Solomon, L. (1977). Zeta functions and integral representation theory. *Advances in Mathematics*, *26*, 306–326. 26.3.14

[Sta67] Stark, H. M. (1967). A complete determination of the complex quadratic fields of class-number one. *Michigan Mathematical Journal*, *14*, 1–27. 25.2.19

[Sta2007] Stark, H. M. (2007). *The Gauss class-number problems*, Analytic number theory: A tribute to Gauss and Dirichlet, Clay Mathematics Proceedings, no. 7 (pp. 247–256). Providence, RI: American Mathematical Society. 25.2.19

[Sun85] Sunada, T. (1985). Riemannian coverings and isospectral manifolds. *Annals of Mathematics (2)*, *121*(1), 169–186. 31.8

[Swa60] Swan, R. G. (1960). Induced representations and projective modules. *Annals of Mathematics (2)*, *71*, 552–578. 20.7.13

[Swa62] Swan, R. G. (1962). Projective modules over group rings and maximal orders. *Annals of Mathematics (2)*, *76*, 55–61. 20.7.13

[Swa80] Swan, R. G. (1980). *Strong approximation and locally free modules*. Ring theory and algebra, III, Lecture Notes in Pure and Applied Mathematics (Vol. 55, pp. 153–223). New York: Dekker. 20.7, 20.7.23, 28.8

[Swi74] Swinnerton-Dyer, H. P. F. (1974). *Analytic theory of abelian varieties*, London Mathematical Society, Lecture Note Series, no. 14. London-New York: Cambridge University Press. 43.4

[Syl1883] Sylvester, J. J. (1883–1884). Lectures on the principles of universal algebra. *American Journal of Mathematics*, *6*(1), 270–286. 1.2

[Tai1890] Tait, P. G. (1890). *An elementary treatise on quaternions* (3rd ed.). Cambridge: Cambridge University Press. 1.1

[Tam66] Tamagawa, T. *Adèles*, Algebraic groups and discontinuous subgroups, Proceedings of Symposia in Pure Mathematics (Vol. 9, pp. 113–121). Providence, RI: American Mathematical Society. 29.11.9

[Tate2010] Tate, J. T. (2010). *Global class field theory*, Algebraic number theory (pp. 163–203, 2nd ed.) J. W. S. Cassels and A. Frölich (Eds.). London: London Mathematical Society. 27.5, 27.5.10, 27.5.13

[Tate67] Tate, J. T. (2010). *Fourier analysis in number fields and Hecke's zeta functions (Thesis, 1950)*, Algebraic number theory (pp. 305–347, 2nd ed.) J. W. S. Cassels and A. Frölich (Eds.). London: London Mathematical Society. 29.2, 29.10.24

[Tho10] Thompson, S. P. (1910). *The Life of Lord Kelvin, Baron Kelvin of Largs* (Vol. II). London: Macmillan. 1.1

[Tit79] Tits, J. (1979). *Reductive groups over local fields*, Automorphic forms, representations and L-functions (Proceedings of Symposia in Pure Mathematics, Oregon State University, Corvallis, Oregon, 1977). Part 1, Proceedings of Symposia in Pure Mathematics (Vol. 33, pp. 29–69). Providence, RI: American Mathematical Society. 23.5.17

[Thu97] Thurston, W. P. (1997). *Three-dimensional geometry and topology*, Silvio Levy (Ed.) (Vol. 1). Princeton: Princeton University Press. 33.8.3, 34.8, 36.5, 36.5.13

[Tig98] Tignol, J.-P. (1998). *Algebras with involution and classical groups*. European Congress of Mathematics, Vol. II (Budapest, 1996). Progress in Mathematics (Vol. 169, pp. 244–258). Basel: Birkhäuser. 4.4.6

[vdB60] van der Blij, F. (1960/1961). History of the octaves. *Simon Stevin*, *34*, 106–125. 1.2

[vdW76] van der Waerden, B. L. (1976). Hamilton's discovery of quaternions. *Mathematics Magazine*, *49*, 227–234. 1.1

[vdW85] van der Waerden, B. L. (1985). *A history of algebra*. Berlin: Springer. 1.2

[vPr2002] van Praag, P. (2002). Quaternions as reflexive skew fields. *Advances in Applied Clifford Algebras*, *12*(2), 235–249. 3.5.3

[Ven22] Venkov, B. A. (1922). On the arithmetic of quaternion algebras. *Izvestiya Akademi Nauk*, *205–220*, 221–246. 1.3, 30.1, 30.2

[Ven29] Venkov, B. A. (1929). On the arithmetic of quaternion algebras. *Izvestiya Akademi Nauk*, *489–509*(532–562), 607–622. 1.3, 30.1, 30.2

[VG74] Vignéras-Guého, M.-F. (1974). Simplification pour les ordres des corps de quaternions totalement définis, C. *R. Acad. Sci. Paris Sér. A*, *279*, 537–540. 32.8

[Vig76a] Vignéras, M.-F. (1976). Invariants numériques des groupes de Hilbert. *Mathematische Annalen*, *224*(3), 189–215. 30.6.18

[Vig76b] Vignéras, M.-F. (1976). Simplification pour les ordres des corps de quaternions totalement définis. *Journal für die reine und angewandte Mathematik*, *286*(287), 257–277. 20.7

[Vig80a] Vignéras, M.-F. (1980). *Arithmétique des algèbres de quaternions*, Lecture Notes in Mathematics (Vol. 800). Berlin: Springer. 21.4.10, 28.7, 28.9.23, 29.3, 30.8.8, 30.9.12, 31.1.8, 31.3, 31.8

[Vig80b] Vignéras, M.-F. (1980). Variétés riemanniennes isospectrales et non isométriques. *Annals of Mathematics (2)*, *112*(1), 21–32.

[Voi2006] Voight, J. (2006). *Computing CM points on Shimura curves arising from cocompact arithmetic triangle groups*, Algorithmic number theory (ANTS VII, Berlin, 2006). F. Hess, S. Pauli, M. Pohst (Eds.), Lecture Notes in Computer Science (Vol. 4076, pp. 406–420). Berlin: Springer. 43.8.2

[Voi2011a] Voight, J. (2011). Characterizing quaternion rings over an arbitrary base. *Journal für die reine und angewandte Mathematik*, *657*, 113–134. 22.2.14, 22.5.13, 22.5.14

[Voi2011b] Voight, J. (2011). Rings of low rank with a standard involution. *Illinois Journal of Mathematics*, *55*(3), 1135–1154. 3.5.5, 3.15

[Voi2013] Voight, J. (2013). *Identifying the matrix ring: Algorithms for quaternion algebras and quadratic forms*, Quadratic and higher degree forms, Developments in Mathematics (Vol. 31, pp. 255–298) . New York: Springer. 9.8

[VW2014] Voight, J., & Willis, J. (2014). *Computing power series expansions of modular forms*, Computations with modular forms, G. Boeckle and G. Wiese (Eds.), Contributions in Mathematical and Computational Sciences (Vol. 6, pp. 331–361). Berlin: Springer. 43.9.7

[VZB2015] Voight, J., & Zureick–Brown, D. (2015). *The canonical ring of a stacky curve*, arXiv:1501.04657. 43.9.6,

[Vos2009] Vostokov, S. V. (2009). The classical reciprocity law for power residues as an analog of the abelian integral theorem. *St. Petersburg Mathematical Journal*, *20*(6), 929–936. 14.6.4

[Wad86] Wadsworth, A. (1986). *Merkurjev's elementary proof of Merkurjev's theorem*, Applications of algebraic K-theory to algebraic geometry and number theory, Part II, (Boulder, Colorado, 1983). Contemporary Mathematics (Vol. 55, pp. 741–776). Providence, RI: American Mathematical Society. 8.3.6

[War89] Warner, S. (1989). *Topological fields*, North-Holland Mathematics Studies (Vol. 157). Amsterdam: North-Holland Publishing Co. 13.5.13

[Was97] Washington, L. C. (1997). *Introduction to cyclotomic fields*, Graduate Texts in Mathematics (Vol. 83, 2nd ed.). New York: Springer. 39.4.11

[Wate69] Waterhouse, W. C. (1969). Abelian varieties over finite fields. *Annales Scientifiques de l'École Normale Supérieure*, 4th series, *2*, 521–560. 42.2, 42.2.16

[Wate79] Waterhouse, W. C. (1979). *Introduction to affine group schemes*, Graduate Texts in Mathematics (Vol. 66). New York-Berlin: Springer.

[Wats62] Watson, G. L. (1962). Transformations of a quadratic form which do not increase the class-number. *Proceedings of the London Mathematical Society (3)*, *12*, 577–587. 25.4.7

[Wats75] Watson, G. L. (1975). One-class genera of positive ternary quadratic forms, II. *Mathematika*, *22*, 1–11. 25.4

[Wed08] Wedderburn, J. H. M. (1908). On hypercomplex numbers. *Proceedings of the London Mathematical Society (2)*, *6*, 77–118. 1.2, 7.3.11

[WW2018] van Weerden, A., & Wepster, S. (2018). A most gossiped about genius: Sir William Rowan Hamilton. *BSHM Bulletin*, *33*(1), 2–20. 1.1

[Weib2013] Weibel, C. A. (2013). *The K-book: An introduction to algebraic K-theory*, Graduate Studies in Mathematics (Vol. 145). Providence, RI: American Mathematical Society. 12.4.4

[Weil60] Weil, A. (1960). Algebras with involution and the classical groups. *Journal of the Indian Mathematical Society*, *24*, 589–623. 4.4.6, 8.4.17

[Weil66] Weil, A. (1964–1966). *Fonction zêta et distributions*, Séminaire Bourbaki (Vol. 9, pp. 523–531). 29.10.24

[Weil74] Weil, A. (1974). *Basic number theory* (3rd ed.). Berlin: Springer. 12.2, 29.3, 29.10.24

[Weil82] Weil, A. (1982). *Adeles and algebraic groups*, Progress in Mathematics. Boston: Birkhäuser. 27.1.11, 27.6, 29.5, 29.8.8, 29.10, 29.10.4, 29.10.24, 29.11.9

[Wein96] Weinstein, A. (1996). Groupoids: Unifying internal and external symmetry. *Notices of the American Mathematical Society*, *45*, 848–859. 19.3.11

[Wes] Weston, T. *Lectures on the Dirichlet class number formula for imaginary quadratic fields*, http://people.math.umass.edu/~weston/oldpapers/cnf.pdf. 25.2.19

[Whe62] Wheeler. L. P. (1962). *Josiah Willard Gibbs: The history of a great mind*. New Haven: Yale University Press. 1.1

[Will2001] Williams, D. P. (2001). *A primer for the Brauer group of a groupoid*, Groupoids in analysis, geometry, and physics (Boulder, CO, 1999). Contemporary Mathematics (Vol. 282, pp. 21–34). Providence, RI: American Mathematical Society. 19.3.11

[Wils2009] Wilson, R. A. (2009). *The finite simple groups*, Graduate Texts in Mathematics. London: Springer. 11.9c

[Zag94] Zagier, D. (1994). *Values of zeta functions and their applications*, First European Congress of Mathematics, Vol. II (Paris, 1992). Progress in Mathematics (Vol. 120, pp. 497–512). Basel: Birkhäuser. 25.2

[Zas72] Zassenhaus, H. (1972). On the units of orders. *Journal of Algebra*, *20*, 368–395. 38.4

Index

© The Editor(s) (if applicable) and The Author(s) 2021

J. Voight, *Quaternion Algebras*, Graduate Texts in Mathematics 288,
https://doi.org/10.1007/978-3-030-56694-4